Technical Report 1118

Team Performance in Distributed Virtual Environments

Michael J. Singer
U.S. Army Research Institute

Stuart Grant
Defence and Civil Institute of Environmental Medicine
Defence Research and Development, Canada

Patrick M. Commarford, Jason P. Kring, and Merrill Zavod
University of Central Florida
Consortium Research Fellows Program

August 2001

United States Army Research Institute for the Behavioral and Social Sciences

**U.S. Army Research Institute
for the Behavioral and Social Sciences**

A Directorate of the U.S. Total Army Personnel Command

**EDGAR M. JOHNSON
Director**

Technical Review by

Dan Dwyer, Naval Air Warfare Center, Training Systems Division
John Barnett, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: TAPC-ARI-PO, 5001 Eisenhower Ave., Alexandria, VA 22333-5600.

FINAL DISPOSITION: This Technical Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE: August 2001	2. REPORT TYPE: Final	3. DATES COVERED: July 1999 – December 2000
4. TITLE AND SUBTITLE: Team Performance in Distributed Virtual Environments		5a. CONTRACTOR OR GRANT NUMBER: N/A
		5b. PROGRAM ELEMENT NUMBER: 2O262785A
6. AUTHOR(S): Michael J. Singer (U.S. Army Research Institute), Stuart C. Grant (DCIEM, DR&D, Canada), Patrick M. Commarford, Jason P. Kring & Merrill Zavod (University of Central Florida)		5c. PROJECT NUMBER: A790
		5d. TASK NUMBER: 202a
		5e. WORK UNIT NUMBER: H01
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: TAPC-ARI-IF 5001 Eisenhower Avenue Alexandria, VA 22333-5600		8. PERFORMING ORGANIZATION REPORT NUMBER:
9. SPONSORING / MONITORING ABENCY NAME(S) AND ADDRESS(ES): U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VE 22333-5600		10. MONITOR ACRONYM: ARI
		11. MONITOR REPORT NUMBER: Technical Report 1118

12. DISTRIBUTION / AVAILABILITY STATEMENT:
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES:

14. ABSTRACT (Maximum 200 words):
The U.S. Army is using virtual simulations for mission planning, training, rehearsal, and concept development. Virtual environment (VE) technology can provide simulated real world activities for dismounted soldiers. One issue in the use of distributed simulations is whether team members learn, perform, and transfer their skills in distributed situations in the same ways as individuals in local situations. In this experiment, local and distributed teams completed a series of mission rehearsals in a VE over two days. Eighteen, two-person teams of college students performed synthetic tasks representative of tasks performed by police, emergency response, and military teams. All participants were trained to criterion in a VE before being assigned to a team. Biographical information and subjective self-report questionnaires were administered before, during, and after training and mission sessions. Local teams interacted face-to-face between mission rehearsal sessions, while distributed teams only interacted by phone during the after action review session following each mission. Local teams performed significantly better than distributed teams on several collective task measures over the repeated missions. Simulator sickness and presence during the mission rehearsals were also investigated.

15. SUBJECT TERMS:
Virtual Environments, Virtual Reality, Team Performance, Distributed Virtual Environments, Distributed Simulation, Communication, Personality, Simulator Sickness, Presence, Immersion, After Action Review

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unclassified	70	Dr. Michael J. Singer, (407) 384-3993

Technical Report 1118

Team Performance in Distributed Virtual Environments

Michael J. Singer
U.S. Army Research Institute

Stuart Grant
Defence and Civil Institute of Environmental Medicine
Defence Research and Development, Canada

Patrick M. Commarford, Jason P. Kring, and Merrill Zavod
University of Central Florida
Consortium Research Fellows Program

Simulator Systems Research Unit
Stephen L. Goldberg, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

August 2001

Army Project Number
20262785A790

Personnel Performance and Training Technology

FOREWORD

The U.S. Army has made a substantial commitment to the use of networked simulations for training, readiness, concept development, and test and evaluation. Many current networked simulators are designed to provide realistic training and rehearsal for large combined arms groups of vehicles and major weapon systems. These simulators represent dismounted soldier activities, but are not intended to directly train or rehearse individual dismounted soldiers. Virtual Environment (VE) technology, which typically includes head-mounted visual displays with tracking devices for limbs and individual weapons, has the potential to provide a more immersive, person-centered simulation and training capability for dismounted soldiers. These systems are being investigated in order to include individual dismounted soldiers in the larger simulation systems, and to support distributed training and rehearsal for teams of dismounted soldiers. One research challenge arising from these efforts is identifying and quantifying the effects of VE system characteristics and use on learning, retention, and transfer of skills required for Army tasks.

This report describes one experiment in an ongoing program of research conducted by the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI), Simulator Systems Research Unit (SSRU) that addresses the use of VE technology for training dismounted soldiers in distributed simulations. This experiment investigated the effects of geographically distributed team members on repeated performance in mission rehearsal exercises. The findings from this research will be used to recommend VE characteristics and instructional methods for incorporation in distributed VE training or rehearsal systems.

SSRU conducts research with the goal of providing information that will improve the effectiveness of training simulators and simulations. The work described here is a part of ARI Research Task 202a, VERITAS - Virtual Environment Research for Infantry Training and Simulation. This work was performed in cooperation with the Defence and Civil Institute of Environmental Medicine, Defence Research and Development, Canada, under the auspices of The Technical Coordination Program, Technical Panel Hum-TP-2, Training Technology Virtual Reality Working Group. The results of this work have been presented to The Technical Coordination Program, Training Technology Technical Panel (Dec, 2000), as well as being presented at several professional conferences.

Zita M. Simutis

ZITA M. SIMUTIS
Technical Director

ACKNOWLEDGMENTS

This research could not have been conducted without the excellent programming and technical support provided by the Institute for Simulation and Training (IST) at the University of Central Florida. The authors especially thank Glenn Martin and Jason Daly of IST for their creative efforts in producing the virtual environment models and data collection software used in the experiment. These data could not have been collected without the assistance and support provided by Robert Wolfe and Don Turner of the Defence and Civil Institute of Environmental Medicine, Defence Research and Development, Canada. Finally, the authors wish to thank Dr. Florian Jentsch of the University of Central Florida for helpful consultation on statistical analyses.

TEAM PERFORMANCE IN DISTRIBUTED VIRTUAL ENVIRONMENTS

EXECUTIVE SUMMARY

Research Requirement:

The U.S. Army is committed to using distributed simulations for mission planning and rehearsal, training, concept development, and test and evaluation. Current systems are designed to provide training for soldiers fighting from vehicles, but are not designed to provide realistic training or rehearsal for dismounted infantry. Virtual environment (VE) technology provides a new way to simulate real world activities for individual dismounted soldiers. This technology may allow the U.S. Army to cost-effectively conduct planning, training, and rehearsal activities for both individual and collective dismounted soldier tasks in distributed simulation systems. Basic to these simulations is the common context of individual combatants who need to move, observe, shoot, and communicate. A key element in distributed systems is whether team members being trained together in geographically distributed situations learn, perform, and transfer their skills in the same ways and at the same levels as team members being trained in the same location. Research on the effects of geographically distributed simulations can establish the benefits, problems, and suggested solutions associated with training and rehearsing complex activities and tasks using distributed VE technology.

Procedure:

In this experiment, 18 two-person teams completed eight mission rehearsals over two days (4 on each day). The intervals between mission rehearsal sessions were no less than one day, and never more than nine days. Nine of the teams were comprised of co-located team members (local teams), while the other nine teams were comprised of one team member in Orlando, Florida, and the other in Toronto, Canada (distributed teams). The local teams met and interacted face-to-face between mission rehearsal sessions, while the distributed teams interacted only by voice (phone) during the after action review (AAR) that followed each mission rehearsal. The tasks performed during the mission rehearsal were synthetic tasks representative of the individual and collective tasks performed by police, emergency response, and military teams in urban (interior) environments. All participants in each condition were trained to perform all tasks and roles to a consistent standard before being assigned to a team. The VE software (identical at each location) enabled collection of task and overall performance data, as well as information about errors. Biographical information was collected, in addition to self-report questionnaires concerning the participants' health, personality characteristics, and reaction to immersive events. Questionnaires were administered before, during, and after the training and mission rehearsal sessions.

Findings:

All teams demonstrated the expected significant improvements in performance on task and collective activity measures over the repeated mission rehearsals. Local teams performed significantly better than distributed teams on several measures of task performance, and maintained that higher level of performance over the repeated mission rehearsals. The primary measures that demonstrated local team superiority were: a) a combined measure of individual and collective task activities involved in conducting an error-free room search, b) the time required to perform the coordinated collective tasks in searching a room (regardless of errors), and c) a measure of loosely coordinated cooperative efforts (coordinated hallway movement). Other, more tightly coordinated collective task measures (door opening and canister disarming), did not show local team performance superiority. Distributed teams were not significantly better than local teams on any measures.

Communication during the AARs, which was coded for communication loops (communications that were verbally responded to by the other team member), revealed no differences between the local and distributed groups. An additional analysis of the communication loops based on a high/low performance split within the groups did not reveal any significant differences between the better and poorer performers. Some indication of personality differences between good and poor performing teams was found in an analysis of team averages on Extraversion, indicating that better performing teams were significantly higher on this personality factor. Analysis of the Simulator Sickness Questionnaire (SSQ, presented in Appendix A) supported prior research by showing a significant decrease in simulator sickness as VE experience accumulated (during training). Analysis of the Presence Questionnaire (PQ, presented in Appendix B) revealed that presence increased as task complexity increased during training, and also increased over the course of the repeated missions.

Utilization of Findings:

The U. S. Army will employ VE technology for training, mission planning and rehearsal, and test and evaluation both in local and distributed formats. Our results indicate that distributed teams may perform more poorly than local teams, which have the opportunity to become familiar with one another outside the learning or mission context. Understanding the possible negative effects of limited involvement between team members in distributed simulation systems will enable developers and trainers to incorporate measures that will avoid those problems. The results of this experiment indicate that there is a need to increase participants' ability to interact in distributed situations, in order to alleviate possible performance degradation during distributed mission rehearsals. The current results do not address longer term skill retention or performance transfer to real situations, nor methods for alleviating differences between local and distributed teams.

TEAM PERFORMANCE IN DISTRIBUTED VIRTUAL ENVIRONMENTS

CONTENTS

Page

List of Tables

Introduction

The U. S. Army is developing programs using many different types of virtual simulation systems for combat training and military concept development, testing, and evaluation (for current information, see the U. S. Army Simulation, Training, and Instrumentation Command [STRICOM] website at www.stricom.army.mil). The early emphasis and implementation of these programs has been on linking vehicle simulators, without providing training for dismounted soldiers (Knerr et al., 1994). The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI), Simulator Systems Research Unit (SSRU), supported by the University of Central Florida Institute for Simulation and Training (IST), has established a research program in Virtual Environment (VE) technology in order to investigate a wide range of potential applications. The program goals are to "improve the Army's capability to provide effective, low cost training for Special Operation Forces and Dismounted Infantry through the use of VE technology and ICS [Individual Combatant Simulation]" (Knerr et al., 1994, pp. 10-12). The program focuses on the VE requirements and application guidelines for leader and individual performance in unit tasks, the determination of necessary characteristics for VE-based ICS training, and the evaluation of transfer of ICS training to military operations.

The original research plan for the overall SSRU program is represented in a hierarchical scheme, the Virtual Environment Research Pyramid (Knerr et al., 1994). The pyramid is based on the military task and activity requirements for dismounted soldier training using VE technology (Jacobs et al., 1994; Levison & Pew, 1993). The lower levels encompass research in psychophysical capabilities required for fundamental soldier activities in VE; the capability in VE of psychomotor acts based on those activities; and comfort, convenience, and side effects in the VE. The middle levels of the research pyramid address the fundamental soldier abilities of spatial knowledge acquisition, terrain appreciation, and route planning in VE, which underlie many soldier activities. The topmost levels of the pyramid focus on studies investigating team leader training using VE, at both the individual and team levels. The research program has never focused on VE-based simulation of a soldier's specific tasks, but has always focused on the fundamental skills that underlie many individual and collective soldier activities and tasks.

Distributed Simulation

Distributed simulation — linked simulations at geographically distant locations — is increasingly being used for military training, concept development, and test and evaluation with individual and small teams of dismounted soldiers. ARI, supported by IST, has established a research program in VE technology, the latest in simulation technology, in order to investigate its application to the training of dismounted soldiers. Similarly, the Defence and Civil Institute of Environmental Medicine (DCIEM), Defence Research and Development Canada (DRDC), is exploring these technologies to extend the benefits of virtual simulation to dismounted combatants. Our groups were brought together by The Technical Cooperation Program (TTCP), Training Technology Technical Panel to investigate joint issues in distributed VE.

Current methods for training and testing dismounted teams on tasks that require interacting directly with the environment are costly and effortful. Typical small unit exercises require gathering soldiers and sending them to a training site (e.g., Project Metropolis exercises; Reeves, 2001). The training site may require extensive development to suit training and rehearsal activities, and cannot easily be altered to present new environmental challenges. Additional challenges are imposed by personnel constraints.

VE systems have the potential to offer effective and less costly alternatives for training and testing dismounted soldiers. VE simulations can support multiple players interacting with computer-generated forces that mimic the behavior of troops, indigenous populations, and enemy forces. VE simulations can also provide multiple simulated terrains and built up areas with appropriate environmental effects, enabling the training to focus on tasks and activities without being limited to unchanging physical arrangements. In addition, VE-based training programs can support a wide range of alterations in the situation, so the team members can practice coordination skills in a number of scenarios and with varying environmental conditions. Finally, performance can be measured with greater ease when training is conducted in a VE.

The VE platform also enables an entirely new type of dismounted soldier team training, one in which the individual team members are physically in different cities, states, or countries, but can still train with one another as if they were in the same locale. However, such a situation may hinder activities that aid in the formation of team cohesiveness. While immersed in a virtual environment, geographically distributed team members are able to see each other's represented body (referred to as an avatar) and movements, and can communicate through the use of microphones and headphones. However, outside the distributed virtual simulation, during an After Action Review (AAR) of their mission performance or other less guided activities, geographically distributed team members may have no communication, or may only be able to communicate over a phone line, with no visual input and little interpersonal feedback. In these situations, because vital interpersonal interactions (e.g. Salas, Dickinson, Converse, & Tannenbaum, 1992) are reduced, relative to geographically local and face-to-face interactions, it is possible that teams performing via distributed virtual simulation will show a decrement in individual and team performance.

Experiment Objectives

This experiment focused on the basic aspect of distributed simulation — the displaced nature of the distributed team — and the possible deficits in team performance or acquisition of skills that might result. As discussed above, during distributed virtual simulation sessions (as well as the associated briefings, reviews, and AARs) team members would not be located in the same physical location. In typical team training and rehearsal, the team members are physically present during prebriefs, rehearsal, and post-activity reviews. In distributed simulations, although all team members are presented with the same information before and after every rehearsal, differences in how team members interact within the distributed situation, both during and between sessions, might change the effectiveness of training. This research is an initial attempt to address this basic aspect of distributed simulation.

In addition, the research was also an experiment in developing a distributed virtual simulation. While there are many instances of distributed vehicular simulations, and even some which include dismounted soldier effects (for example, the Close Combat Tactical Trainer [CCTT] program; see http://www.stricom.army.mil/STRICOM/PM-CATT/CCTT/), at this time (to our knowledge) there are no geographically distributed simulation networks exclusively for dismounted soldiers (or even any that employ a high percentage of autonomously interacting individuals). As such, at the outset, we were not sure that the experiment would be possible. The fact that the system did work in a robust fashion is due to the technical expertise of the programmers resident at IST (under contract to ARI, SSRU).

The primary psychological objective of the present experiment was to investigate whether teams whose members complete a series of simulated rehearsals and AARs in the same physical location would perform differently than teams whose members complete rehearsals and AARs remotely, with more restricted, non-rehearsal interactions. We decided that the framework for the team missions should be generic, with activities that represented a wide range of individual and collective tasks. To achieve this objective, as well as better understand the use of VE technology for team training or rehearsal in general, this experiment evaluated several wide-ranging factors that have the potential to influence individual and team performance in the virtual task. The following sections outline the main characteristics of teams and team training, followed by variables that might affect team performance including: communication and personality. In order to clearly frame the factors reviewed, we first present an outline of the nature of the team mission used in the VE mission, and the overall structure of the experiment. Further details are provided, as usual, in the Methods section. During the course of the research, as is common in the SSRU VE research program, we also investigated simulator sickness and immersion and presence (as experienced in the VE). The material on simulator sickness is presented in Appendix A and the material on presence is covered in Appendix B.

The present experiment employed a set of synthetic tasks based in multi-room building environments that would provide face validity for the participants, and enable generalization of results to other environments and training situations. Each participant initially undertook individual training in all basic skills required for <u>both</u> team roles and those general skills required by the VE equipment configuration, during a training phase. After training, participants were assigned to a team and began repeated mission rehearsals. The teams consisted of two participants, each performing both common tasks and role-specific individual tasks. The repeated mission rehearsals provided a learning background that set the context for possible differences in team member location. As the team progressed through the successive mission rehearsals, performance on the individual and collective tasks would naturally improve. Any differences in team or individual performance between the differently composed teams (both local or distributed at different locations) could be attributed to the composition. An AAR was administered after each mission rehearsal to provide feedback on team performance. The basic timeline and experiment design is presented in Table 1.

Table 1
Experiment Phases

Training Phase		Mission Rehearsal Phase	
Individual Session (4 h)s.	Team Assignment	Session One (4 hrs.)	Session Two (4 hrs.)
Movement Communications Equipment Use Team Tasks	Local Team	Both Team Members at SSRU	
	Distributed Team	One Team Member at SSRU, One Team Member at DCIEM	

(All sessions on different days.)

Teams and Team Training

A number of definitions and models of what is a team, and what distinguishes teams from a mere collection of individuals, have been proposed by training researchers. For example, teams are generally considered to be different from groups, mobs, or collections of individuals. Perhaps the most encompassing description of a team is provided by Salas et al. (1992): two or more individuals with a common goal that requires coordinated, interdependent, and adaptive performance. This broad definition implies that there are many widely ranging and interacting factors that can affect team performance. The common team goal requires that a set of individual and collective tasks be performed during a specific time frame. The nature of the tasks dictates the required resources, individual skills, and team member interdependence. As the task-required interdependence increases, communication and understanding between members becomes more crucial in achieving the group's goals. A plethora of additional factors and dimensions can be examined in discussing group behavior and performance, depending upon the goals and level of analysis. Such dimensions and factors include the individual unit versus the team unit, personality factors and skill levels of the team members (individually or in some concatenation), the structure of the group, the place of the group within a larger organization, the life cycle of the group, and so forth. These concepts suggest that there are many factors that could be affected by a team's distribution and the decrease in communication capability.

Hackman (1993) identified many salient factors of team performance as key elements for team effectiveness: ability to work together, satisfaction of member needs, acceptability of outcomes, level of effort of members, individual skill and knowledge levels, task appropriateness, and resource allocation. Gersick (1988), and the Team Evolution and Maturation (TEAM) model developed by Morgan, Salas, and Glickman (1993), on the other hand, focused on identifiable patterns in the lifecycle of the team and its individuals from time of formation to the dissolution of the team. Many of these concerns have been revisited by training researchers addressing teamwork or performance. In an experiment, the participants know that they are only brought together for the duration of the work. That may change the response patterns, so that the experimental participants respond in different ways than professionals that are practicing and rehearsing vital job skills and team routines. There is no way to know this without conducting a relatively clever (and potentially impossible) experiment. In lieu of

conducting that effort, one must assume that the participants are approaching the team tasks in a serious fashion, much as professionals would. With that as a background assumption, our efforts to produce local and distributed teams proceeded.

Research requires decisions on the type and level of analysis required for evaluating team training and rehearsal. Teams are composed of individuals, and this reality carries with it several implications. First, the patterns of communication and ability of the individuals to cooperate with one another affect the team as a whole. For instance, Stout, Salas, and Carson (1994) showed that team interaction and coordination was associated with mission performance for 2-person pilot teams involved in a low-fidelity flight simulation. Similarly, Bowers, Jentsch, Salas, and Braun (1998) found that more successful teams communicate significantly more with one another than unsuccessful teams during task performance. This implies that without adequate individual communication skills, the task performance capabilities of the team are limited. Communication is presumably based on and supports a shared mental model of the situation state and tasks, so that team members are able to work together as opposed to operating at cross purposes. These issues are discussed in more detail in the communication section that follows. Second, the skills of the individual influence team performance. Researchers have shown that individual cognitive ability and job-related skills are related to team performance (Comrey, 1953; Terborg, Castore, & DeNinno, 1976). For example, Terborg et al. (1976) found that during a land surveying task, teams with members possessing high cognitive ability performed better than teams comprised of members with lower cognitive abilities. It therefore appears that the general cognitive abilities of the individual members of a team are reflected in overall team performance.

Based on the research reviewed above, it seems reasonable to hypothesize that the local teams will perform better overall than the distributed teams. Although the mission-based interactions will be equivalent for both local and distributed groups, the local teams ability to interact face to face will probably ease the team feedback and formation that is needed for success in collaborative tasks. Second, it is obvious that all teams will improve significantly over the repeated missions. Humans will learn and improve quickly in most situations (unless task difficulty is great, and our tasks are designed not to be extremely difficult). The main reason for studying repeated missions is to investigate possible interactions. There is no research evidence that would indicate whether our hypothesized local team advantage will decrease, increase, or remain constant over missions.

This discussion about teams and the influences on team performance raises the issue of measurement. Both individual and collective tasks in the military are typically measured in terms of go/no go criteria applied by subject matter experts (SME). This provides a gross measure that assures that tasks can be adequately performed, and may not take into account the individual contributions by team members (Tesluk, Mathieu, Zaccaro, & Marks, 1997). Overall outcome measures of team success (e.g., mission success) may be appropriate if the team behaviors contributing to mission success are intensive. Outcome measures can indicate team success, but process measures can provide information for team improvement (Johnston, Smith-

Jentsch, & Cannon-Bowers, 1997). Multiple measurements were used in the research in order to obtain the most complete picture possible of the team's performance and mission success.

Factors Moderating Team Performance

A number of factors have the potential to influence individual and team performance in virtual tasks. For this reason, the present experiment assessed several characteristics related to the team tasks. In addition, for research with state-of-the-art technologies and complex interactions between people, it only seems reasonable to collect as much relevant information as possible. SSRU has a long history of research on and about VE systems, and in virtually all of that research we have employed multiple measures and questionnaires addressing not only the direct issues framing the research, but also general issues in VE use. Given the long history of sickness associated with exposure to simulators, information is gathered to ensure that no harm comes to participants in our research (see Appendix A). As the research program is designed to provide general information and knowledge about VE use, materials that address the participants' responses to the VE experimental situation are also used (see Appendix B).

The following subsections present additional factors addressed in the present experiment. The sections address communications and personality factors that can help explain the results of the research. The subsections provide a brief review or background as a basis for inclusion in the research, an exposition of the method used for addressing the factor in the context of the current research, and hypotheses about the outcome of the current research relevant to the additional factors.

Communication. Effective team training has long been a goal of the military as well as other organizations concerned with maximizing team performance. Effective training requires an understanding of team processes and identification of specific behaviors that can maximize team productivity and minimize errors. It is generally believed that the use of appropriate communications, both during and between tasks, can greatly improve performance in a variety of disciplines (e.g., Jentsch, Sellin-Wolters, Bowers, & Salas, 1995). The analysis of team communication styles and how these styles relate to performance is an area of research that is increasingly being explored.

Several research efforts have examined the relation between performance and amount of communication during team activities with mixed findings (Jentsch et al., 1995; Mosier & Chidester, 1991). The results indicate that identifying specific patterns of communication that are most conducive to team success appears to be a more promising endeavor. Kanki, Lozito, and Foushee (1989) found that air crews using speech that is consistent in content and in speaker sequence during flight operations outperform teams whose speech is absent of these qualities. Building on Kanki et al.'s (1989) work, Bowers et al. (1998), in a series of two studies, revealed several patterns of communication that were indicative of better performing teams during simulated flight tasks. They demonstrated that an analysis of two-statement communication sequences discriminated between good and poor teams to a much greater degree than simple communication frequency counts. Bowers et al. (1998) found that poor teams closed a lower

proportion of total communication utterances with responses (as opposed to leaving the loop open, characterized by no response or an irrelevant response from the team member after an utterance) than good teams. Poor teams specifically followed a lower proportion of facts, planning statements, uncertainty statements, and action statements with acknowledgements. These poorer-performing teams also used a higher proportion of non-task related communications, were less likely to follow action statements with other action statements, and were less likely to follow communication from air traffic control with planning statements. It should be apparent from the last point that these communication sequences were collected during the simulated flight task. Both Kanki et al.'s (1989) research and Bowers et al.'s (1998) work focused solely on the relationship between in-task communication patterns and team performance.

Research examining between mission communication patterns and team performance is limited. Between mission, or intermission, communication covers the discussion between team members during non-task or mission activities, for example crew discussions before or between flights rather than during actual flight activities. Some research has focused on planning behaviors, measured in terms of communication, which has been shown to relate to team performance. For example, Orasanu (1990) found that better-performing teams used more planning, especially in times of low-workload during their activities. Alternatively, Stout, Cannon-Bowers, Salas, and Milanovich (1999) asked raters to evaluate the quality of planning between teammates during a pre-mission communication session, and found that this measure was related to subsequent in-mission performance. Their work indicates that perceived effective team planning can escalates team performance, possibly based on shared mental models among teammates, which in turn improves team communication during conditions of high-workload. In contrast, Meliza, Bessemer, & Hiller (1994) discussed appropriate methods of administering an intermission AAR for the purpose of maximizing future team performance in a distributed simulation setting.

The present experiment focused, in part, on identifying the relationship between AAR communication patterns and subsequent performance. The experiment also sought to determine if significant differences exist between the natural communications of local teams (who have face-to-face communication capabilities) and distributed teams (who have only voice communication) during AAR sessions. This kind of information might help us better understand the relationship between locality and performance, in the absence of interventions like directed planning or training in team coordination. The content categories used by Bowers et al. (1998) were developed specifically for the coding of in-mission communication by flight crews, thus some changes were made to develop appropriate scoring for our purposes. Based on our review of the literature, we expected to find differences in the communication patterns of good and poor teams, but there was no overt reason to expect differences in communication between local and distributed teams. Specifically, our hypothesis was that better-performing teams would have higher levels of communication on a number of measures. These measures include the percentage of utterances with responses, the number of planning statements, the proportion of planning utterances, the proportion of non-mission related utterances, the proportion of mission-related questions, and the proportion of planning statements.

Personality. The present effort also provided an opportunity to address the effect of personality on team performance in the virtual task. Personality, like other team member characteristics (e.g., skill level, communication patterns, motivation, resource allocation, workload, task behaviors; Salas et al., 1992) has been shown to be a reliable predictor of performance in team tasks (e.g., Jackson, 1992; Moreland & Levine, 1992; Neuman & Wright, 1999). Furthermore, as the interaction of two or more individuals is a central feature in team performance, team members' capacity to attend to input from others in an interdependent fashion is crucial to overall performance. Driskell and Salas (1992) term this capacity *collective behavior*, referring to "...the tendency to coordinate, evaluate, and utilize task inputs from other group members in an interdependent manner in performing a group task" (p. 278). Because an individual's personality guides how he or she interacts with others, we surmised that personality traits could have an important influence on collective behavior, and thus overall team performance.

Personality, defined as stable, deep-seated predispositions to respond or behave in particular ways that are relatively consistent over time and across situations (Chidester, Helmreich, Gregorich, & Geis, 1991), has received extensive attention in the human factors and industrial/organizational psychology literature. Some recent studies of personality and individual and team performance were inconclusive (Driskell, Hogan, & Salas, 1988), due in part to the lack of a common definition of personality (Neuman, Wagner, & Christiansen, 1999), the use of divergent personality measures in research (Chidester et al., 1991), and the absence of a standard framework to organize measures and empirical results (Aiken, 1989). To overcome these obstacles, researchers have more recently employed a five-factor model (FFM), or "Big Five" theory of personality, which categorizes a multitude of personality traits into five primary domains: Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. These five domains, and their respective descriptions and representative adjectives, are presented in Table 2 (adapted from Costa & McRae, 1992; Vickers, 1995).

Research based on the FFM has provided support for a personality-performance relationship. At the individual level, personality has predicted performance in Army personnel (McHenry, Hough, Toquam, Hanson, & Ashworth, 1990); health care and service employees (Rosse, Miller, & Barnes, 1991); the leadership abilities of military academy leaders (Atwater & Yammarino, 1993); U.S. Coast Guard Academy graduates (Blake, Potter, & Slimak, 1993), and U.S. Naval Academy graduates (Atwater, 1992). In a majority of these studies, Conscientiousness is the personality factor most strongly, and consistently, associated with individual performance (Barrick, Stewart, Neubert, & Mount, 1998; Bing & Lounsbury, 2000).

Personality has also been shown to predict performance at the group level. The personality styles of leaders in team and group situations, for example, have been correlated with overall team or group performance (e.g., Atwater, 1992; Atwater & Yammarino, 1993; Chidester & Foushee, 1989). Vickers (1995) reviewed an experiment by Blake, Potter, and Slimak (1993)

Table 2
Descriptions of the Big Five Personality Factors

Domain	Description and Representative Adjectives
Neuroticism	High scorers tend to express negative affects like fear, sadness, anger, guilt, disgust, are more susceptible to psychological distress, more prone to have irrational ideas, are less able to control impulses, and cope poorly with stress. Representative adjectives for high scorers include: ♦ Anxious, fearful, worrying, irritable, impatient, excitable, high-strung, pessimistic, hasty, temperamental, sarcastic, envious, insecure.
Extraversion	High scorers tend to be sociable and exhibit more upbeat, optimistic attitudes. High extraverts also talk more and enjoy excitement and stimulation. Adjectives include: ♦ Friendly, warm, cheerful, social, outgoing, aggressive, assertive, forceful, enthusiastic, energetic, determined, active, daring, adventurous, spontaneous, humorous.
Openness to Experience	High scorers are generally more curious about, and attentive to, their inner world or experience as well as the external environment than low scorers. High scorers also have active imaginations, greater aesthetic sensitivity, preference for variety, independence of judgement, and a willingness to entertain novel ideas and unconventional values. High scorers also experience positive and negative emotions more keenly than low scorers. Adjectives for high scorers include: ♦ Imaginative, idealistic, intellectual, curious, artistic, original, inventive, unconventional, complex, deep.
Agreeableness	High scorers are generally more altruistic and sympathetic toward others than low scorers. High scorers also believe others will be equally helpful in return. In contrast, low scorers tend to be antagonistic, egocentric, skeptical of others intentions, and competitive rather than cooperative. Adjectives for high scorers include: ♦ Warm, gentle, kind, considerate, sympathetic, helpful, generous, tolerant, trusting, forgiving.
Conscientiousness	Factor is related to one's ability to resist impulses and temptations as well as the ability to plan, organize, and carry out tasks. High scorers are generally purposeful, strong-willed, punctual, and determined. Further, high scorers tend to exhibit greater achievement motivation in academic and occupational settings. Adjectives for high scorers include: ♦ Ambitious, industrious, efficient, determined, persistent, prompt, thorough, organized, precise, methodical, resourceful, self-confident.

Note. Adapted from Costa & McRae, 1992; Vickers, 1995.

that compared personality measures and performance ratings of U.S. Coast Guard Academy graduates. Higher leadership ratings, based on established officers' overall rating of each cadet, were given to graduates with high levels of Extraversion and Conscientiousness and low levels of Neuroticism.

Concerning team performance specifically, Barrick et al. (1998) assessed performance and viability— the capability of team members to continue working together cooperatively—for work teams in a manufacturing facility. Results indicated that teams with higher mean Conscientiousness levels received higher supervisor ratings for team performance than teams with lower mean Conscientiousness levels. The authors partially attributed this finding to the fact that achievement motivation is a component of the Conscientiousness factor, and that teams with members exhibiting high achievement motivation generally perform better than low achievement motivation teams (e.g., French, 1958; Schneider & Delaney, 1972). In addition, Barrick et al. showed teams with higher mean levels of Extraversion and emotional stability (i.e., low Neuroticism) received higher viability scores. In a similar study, Neuman and Wright (1999) found that Conscientiousness and Agreeableness were positively correlated with task performance, at the individual and group level, for four-person human resource teams. Conscientiousness was also related to team performance in a study of mixed-gender teams (Kickul & Neuman, 2000). Based on the above literature, our first hypothesis is that high-performing teams will exhibit higher mean levels of Conscientiousness than low-performing teams. We also expect that high-performing teams will exhibit lower mean levels of Neuroticism than the low-performing teams.

Two of the other personality factors also appear to be related to team performance: Agreeableness and Extraversion. Costa and McRae (1992) proposed that these two factors are primarily dimensions of interpersonal tendencies. Because the team tasks in the VE were structured to require cooperation, communication, and team interaction, we predicted that teams with members proficient at interpersonal interactions, as should be evidenced by high Agreeableness and Extraversion scores, would perform at a higher level. In other words, high-performing teams will exhibit higher mean levels of Agreeableness and Extraversion than low-performing teams.

Findings also indicate that the pattern or mixture of personality variables, not just mean levels of each variable, affects team performance. Neuman et al. (1999), for example, analyzed the relationship between team effectiveness and personality in teams of retail personnel. Average levels of Conscientiousness, Agreeableness, and Openness to Experience were positively related to team performance, consistent with other research. However, dissimilarity in Extraversion and Neuroticism were also positively related to team performance. Teams with diverse levels of these factors (e.g., some members high, some members low) exhibited better team performance. Neuman et al. argued this team heterogeneity, or team personality diversity (TPD) improves performance because "...each member adds unique attributes that are necessary for the team to be successful. For example, a team that is heterogeneous with respect to Extraversion may perform effectively because some members fill the role of being outgoing and leading, whereas others fill the role of being reserved and following" (p. 31). Based on Neuman

et al.'s findings, and the fact that a leader-follower dimension was used in their research, similar to our own (i.e., the Team Leader/Equipment Specialist roles in the present experiment), another hypothesis was developed for the mixture of personality factors present in a team. The high-performing teams will exhibit more diverse levels (e.g., one team member high, the other low) of Extraversion and Neuroticism than low-performing teams. Note that this prediction contrasts with the previous hypotheses on the overall levels of Extraversion and Neuroticism in the teams.

Training and After Action Review

One ubiquitous training technique in the military is the AAR (Brown, Nordyke, Gerlock, Begley, & Meliza, 1998). This classic and basic learning principle is often referred to as "knowledge of results" or feedback in the general literature, and used in many different ways as an instructional technique (e.g., Goldstein, 1974). The military uses this technique to review the decision points, key situational factors, and other actions made during an exercise. During the present program, the AAR is used to review the activities performed during the mission, and correct or improve performance speed and accuracy. In the AAR, participants learn how well they did, examine exactly where mission processes were not optimal, and review situations in order to identify problems in timing, procedure, and planning. A review of critical sequences in the mission can also help identify cueing stimuli that may have been missed or used inappropriately. The AAR is not the focus of this research, but is used throughout the work to provide the opportunity for team members to analyze and attempt to improve on their accuracy or speed of performance. The framework for their efforts is presented in the Methods section.

Methods

Participants

Participants were acquired from two geographical locations: Orlando, FL and Toronto, Canada. All participants had normal or corrected-to-normal vision and no significant physical health problems. Only a portion of the total number of participants were actually assigned to a team for the mission phase because some participants either (a) did not meet minimal training requirements (i.e., did not achieve criterion), (b) could not return for the mission phase, or (c) dropped out of the experiment due to simulator sickness or other complications. Subsequently, the training participant sample and the team participant sample are described separately.

Training Participants. Participants ($N = 64$) were drawn from two locations. Orlando participants were students (40 men and 14 women, median age = 20 years) from the University of Central Florida. Participants in Toronto, Canada were co-op students (9 men and 1 woman, median age = 22 years) from a number of universities that were working at DCIEM.

Team Participants. Team participants ($N = 36$) were a subset of the trained participants, as discussed above. Twenty-seven team participants (21 men and 6 women, median age = 21 years) were from Orlando, and nine team participants were from Toronto, (8 men and 1 woman, median age = 22 years). All mission rehearsal (teamed) participants had successfully completed

training, and had schedules that enabled a team to be formed relatively soon after individual training. No team began missions more than a week after training, and no teams were formed with female participants in both roles. (We did not want to limit teams to only males, as females do participate in different kinds of distributed virtual simulations. However, we did not want a dramatic imbalance in team makeup that could not be analyzed. Therefore, we attempted to balance the distribution of the sexes, with the caveat of not having an all female team.) Participants were assigned to either local or distributed teams, with the local (Orlando only) teams having 14 males and 4 females (median age = 20.5) and the distributed (Orlando and Toronto) teams having 15 males and 3 females (median age = 22). These pairings produced nine local and nine distributed teams.

Materials and Equipment

Questionnaires. Questionnaire information was collected using an Access™ database developed by ARI researchers, implemented on a standard Windows95™ platform. Four questionnaires were used and all were presented via the Access™ program. Hard copies of the questions are contained in Appendices C-F. The biographical questionnaire addressed basic demographic statistics, health, motion sickness history, and computer, video, and virtual reality gaming experience and use (Appendix C). Additional questionnaires were the ITQ (Appendix D), which addresses tendencies toward involvement in experiences, the PQ (Appendix E), which assesses immersion and involvement aspects of the immediately preceding experience, and the SSQ (Appendix F), which assesses simulator sickness symptoms. Both the SSQ and PQ were administered repeatedly throughout the experiment as described in the procedures and discussed in Appendices A and B, respectively.

The five personality factors were assessed with the NEO Five-Factor Inventory (NEO-FFI™, Costa & McRae, 1992; copyright by Psychological Assessment Resources, Inc.), a shorter version of the NEO Personality Inventory (NEO-PI™; Costa & McRae, 1992; copyright by Psychological Assessment Resources, Inc). The NEO-FFI provides estimates of the five personality factors based on participants' responses to a series of 60, Likert-scale questions. Raw scores are transformed into *T* scores for easy comparison to norms in the general population and the development of personality profiles.

Virtual Environment. The VE was rendered at both sites on Silicon Graphics Onyx™ computers with Reality Engine graphics sub-systems. MotionStar™ sensors were used to track participant's physical movements, and Virtual Reality VR8 head mounted displays (HMD) were used to present head-slaved computer-generated, stereoscopic color imagery to the participants. Stereo sound was provided through earphones attached to the HMD. The sound included voice communications between each of the participants and the experimenter, and sound effects that included collision noises, doors opening, grenade explosions, and gunfire. The software was written by IST using Performer, C++, and Java.

Mission Rehearsal VE. As described above, the mission rehearsal scenarios were ten-room building VEs laid out along a single corridor, scaled approximately four meters wide with

one ninety degree turn, either to the right or left. The buildings were designed to represent normal offices, a school, a department store, a library, a warehouse, and single story homes (see Figure 1 for an example layout). The corridors were all scaled to 70 meters in length, with the

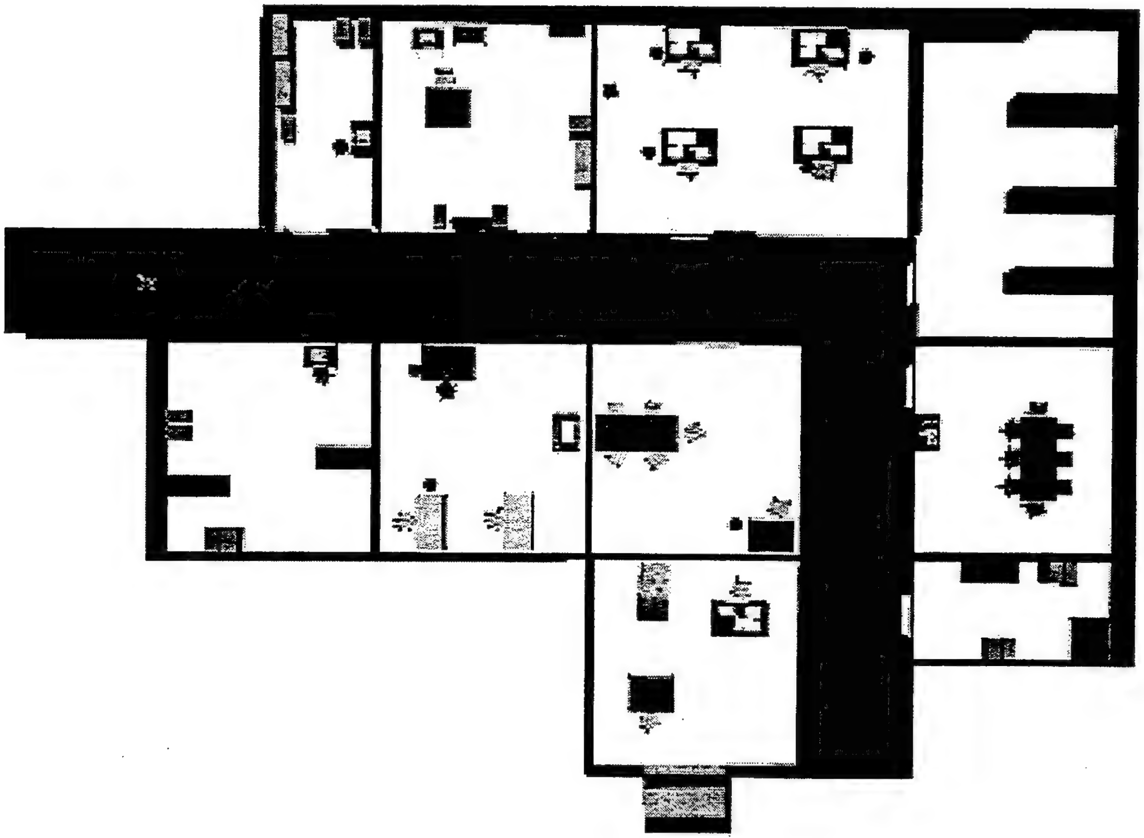

Figure 1. Example Layout of Mission Scenario

turn at 20, 25, or 30 meters. The rooms varied between 5 x 10 and 15 x 10 meters in size, with scenarios being furnished in themes; office furniture, home furnishings, warehouse shelves, library bookcases, retail store appliances and furnishings, or classroom desks. The rooms in Figure 1 represent the office theme, with a small library in the room on the top right corner of the figure, and offices with desks, tables, and chairs in the other rooms. Teams would enter from the small room at the bottom, as if a van had backed up to the door into the building. This eliminated any activities outside the building.

The scenarios were populated with varying numbers of Neutrals (VE avatars that have no weapons) and opposing forces (OpFor, VE avatars that are holding and using weapons). Avatars were all human forms that had normal civilian appearances, so that the only discriminating factor between Neutrals and OpFor was whether the avatar was holding a weapon and firing on the team. All scenarios also had varying numbers of gas canisters, which also varied in their

placement and state. Canisters had one of three possible armed states: a) no gas & not armed, b) gas & not armed, and c) gas & armed. The participants were instructed that the gas in the canisters was harmful for civilians, but not for team members, as they were wearing Hazardous Materials (HazMat) suits. Scenario complexity (based on the number of OpFor, and the number and state of canisters) was balanced across the different scenarios to the greatest extent possible. For example, with several armed gas canisters and a few unarmed canisters per scenario, the canisters could not appear in every room in every scenario, nor in the same order of room in each scenario. Typically an armed canister was encountered in at least one of the first three rooms. The order of scenarios was randomized such that each team received a unique permutation of scenarios, and across teams, no scenario was first or last more than once. Teams were instructed not to proceed past the X on the floor at the end of the corridor, which effectively limited the area for the mission (see Figure 1).

Networking. VE system data were exchanged between the local computer networks for the local teams, and the networks were connected over an ISDN line for the distributed teams. Voice communications between the players during both the mission rehearsals and the AARs were carried on commercial telephone lines.

Procedures

The SSRU in Orlando had a larger participant pool from which to draw, and conducted all of the local mission rehearsals and data collection, as well as training half of the participants for the distributed teams. All participants were kept unaware of the distributed team focus of the research. During briefings, the learning aspects of the repeated mission trials were repeatedly emphasized.

Orlando participants received monetary compensation for all time spent in training and mission rehearsals, with bonuses provided for completing training and returning for all mission sessions. Toronto participants volunteered as a part of their internships, and were not further compensated. All local teams completed mission rehearsals at the same location: a laboratory at IST in Orlando, FL. For distributed teams, one participant was located in Toronto at the DCIEM laboratory, and the other participant was at IST.

Participants were first informed about the general nature and requirements of the VE and training and mission rehearsals. This introduction included viewing a video that demonstrated the VE equipment, special techniques for using the equipment, and mission tasks. Participants were also told about the multi-session nature of the experiment in order to ensure each was committed to multiple sessions. Following the introduction, participants gave consent to participate and then completed a biographical questionnaire, the ITQ, and the initial SSQ before starting the training program.

Training. Training occurred at both the Orlando and Toronto locations. During a single training session, which averaged 4 hours, each participant learned communication protocols and how to perform the primary tasks required in the mission rehearsals (e.g., walking, door opening,

grenade launching, gas canister detection and disarming). This was done by having participants first watch a demonstration of the task, and then practice the task with the experimenter (for communication protocols) or in the VE (for physical tasks). The training concluded with practice on the major coordinated team activities with an automated partner in the VE. All participants were trained to perform both roles: team leader (TL) and equipment specialist (ES). As noted above, each role had specific duties within the mission context. Furthermore, all participants were required to reach a predetermined criterion of no significant errors on any task in order to be assigned to teams for the mission rehearsals. Errors in a task required the participant to repeat the task until achieving acceptable performance.

All training was completed at least one day prior to the first session of team mission rehearsals. During the experiment, in order to minimize any adverse effects of immersion in the VE, participants were only allowed to spend a maximum of 12 accumulated minutes immersed in the environment within a 30-minute time frame (the 30 minutes started at initial exposure to the VE). The participants then had a minimum 30-minute recovery time between VE immersions, during which questionnaires and non-VE training were administered. After the first VE training session, which trained movement using the virtual environment equipment, participants completed another SSQ and their first PQ. Subsequently, an SSQ was administered before and after every VE session, and was also administered 30 minutes after the last VE session of every day (see Appendix A for analyses, results, and discussion). This ensured that an evaluation of symptoms was completed before the participant was released for the day. If symptoms were elevated, the participant was kept on-site until symptoms diminished to near normal. The PQ was also administered again immediately after the last VE training session (see Appendix B for presence and immersion analyses, results, and discussion).

Mission Rehearsals. Following training to criterion, each participant was randomly assigned to a team (local or distributed) using counterbalanced assignment of team roles within the distributed group. Once assigned to a team, the participant did not change their role or teammate during the mission rehearsal trials. Each team completed two sessions during which 8 mission rehearsals were performed (four during each session). The two sessions occurred on separate days, with a minimum of one day, and a maximum of seven days, between sessions. In each mission rehearsal the team moved through one of the ten-room building scenarios, searching for and disarming gas canisters, dealing with OpFor and neutrals, as described above.

As with the training, in order to minimize any adverse effects of immersion in the VE, participants were only allowed to spend a maximum of 12 to 15 accumulated minutes immersed in the environment within a 30-minute time frame (the 30 minutes starting at initial exposure to the VE). This exposure limitation was selected based on our experience with VE systems and earlier research showing that simulator sickness increases with increasing exposure times. The exposure limitation was accomplished by having the team begin their exit from the scenario at the ten-minute mark after VE initiation, and the VE would automatically freeze after twelve minutes in the mission. This time did not include the initial equipment check and alignment routine at the start of the missions.

After each mission rehearsal, the participants had a minimum 30-minute recovery period before the next mission rehearsal, during which questionnaires were administered. As during the training program, an SSQ was administered before and after every VE session, and was also administered 30 minutes after the last VE exposure of each day (see Appendix A). This ensured that an evaluation of symptoms was completed before the participant was released for the day. If symptoms were elevated, the participant was kept on-site until symptoms diminished to near normal. The PQ was also administered again immediately after the first and last mission rehearsals (see Appendix B for analyses, results, and discussion).

After Action Review. At the conclusion of each mission rehearsal, the team conducted a 10-minute AAR. The experimenter at IST in Orlando acted as a reviewer, replaying two critical segments of the mission rehearsal for which performance was sub-optimal. Each AAR was broken down into two separate five-minute segments: the first focused on the mission protocol (accuracy emphasized), and the second on mission performance speed. The mission segments were selected for replay based on a pre-established hierarchy of errors (with the most complex collective tasks ranked as most important and search patterns and movement rated as least important). The segment with the most critical error was then selected for review. During the AAR, the experimenter provided a written example of the correct protocol for each segment (a room search or hallway movement activity), and participants were instructed to discuss what happened, why it happened that way, and how they could improve performance during the next mission. During the AAR period, after the team completed their desired discussion, they were allowed to address other aspects of the mission in which they perceived problems.

In the local condition (at IST Labs near SSRU in Orlando), team members communicated face-to-face with one another and the reviewer during the AAR. In addition, after completion of the AAR, local team members were allowed to communicate with each other on an interpersonal level concerning non-mission topics. Participants were instructed not to discuss mission topics during these free periods, and were admonished when caught discussing techniques or activities (which seldom happened). These free intervals were often limited by the requirement to fill out questionnaires during the recovery interval between VE missions and typically varied from a few minutes to as much as fifteen minutes.

In the distributed condition, the reviewer/experiment controller was in the same room as one team member (at IST), but the other team member was located at DCIEM (in Toronto). In this condition, the team members communicated only by voice (over phone lines, see above) and only during the AAR replay (presented simultaneously at each location). The AAR was conducted in as near an identical manner to the local team AAR as possible (given the need to verify communication and time the start and end of the replay at each site). Distributed team members did not have an opportunity for any interpersonal discussion after the AAR, although occasional interpersonal comments did occur during the AAR period, after the team had completed their desired discussion of the mission.

Results

As discussed above, we trained a larger number of participants than were actually assigned to teams and completed the mission rehearsal phase of the experiment. The training and training-associated questionnaire results cover all participants who successfully completed training ($N = 64$). The results from the team data are presented separately ($N = 36$, 18 Teams), along with team or individual measures and the questionnaires associated with mission rehearsals. The training information and questionnaire results are presented first, the team mission performance results are presented afterward, and the section ends with associations between mission rehearsal performance and the questionnaire results collected during the mission rehearsals.

Training

VE Trials. The number of VE trials required for training was the most reliable data available for analysis of possible differences in training. The first analysis is between overall training at SSRU (Orlando) and DCIEM (Toronto). SSRU trained all local participants and half of the participants later assigned as distributed team members, for a total of 54 participants (27 of whom were used for the local and distributed teams), while DCIEM trained 10 participants (nine of whom were used for the distributed teams). The results of a planned comparison *t*-test on the overall number of VE sessions administered during training found a significant difference in the average number of VE sessions used during training between the locations (t (adj. df 10.227) = 2.887, $p = .016$). Adjusted degrees of freedom were used because the number of participants at each location (and the variance of the groups) was so different. The mean number of sessions for the SSRU trained participants was 3.1667 while the DCIEM trained participants averaged 2.5 sessions in VE training.

Mission Rehearsals

Team performance was measured in a number of ways, as indicated in the introduction. The more complex individual and collective tasks were considered most likely to provide clean evidence about any possible differences due to team member location. The measures used addressed the team's time to complete cooperative tasks and activities, correctness and timing of task interactions, and the overall accuracy of collective task performance.

Training. Since there were apparent differences in at least the number of VE sessions used during overall training at the different locations, we inspected the training information for the participants that formed the teams used in the experiment. A test of the number of VE training sessions administered for the SSRU participants versus the DCIEM participants found a significant difference (t(adj. df 8.731)=2.543, p=.032; SSRU trained = 3.07, DCIEM trained = 2.44). As before, the adjusted degrees of freedom were used to compensate for the unequal number of participants trained at each location. The distributed teams used in the experiment were comprised of both SSRU and DCIEM trained participants, however. A comparison of the mean training sessions for the participants comprising the actual teams used during the

experiment found no significant differences in training between the local and distributed teams. The means for those teams are local equal to 3.06 and distributed equal to 2.78. Finally, an analysis of the difference in performance on the number of rooms searched correctly and successfully during the initial mission rehearsal session was conducted (see below for a description and the overall analysis of this variable). The analysis did not find a significant difference in performance between the local and distributed teams during the first mission rehearsal. These results indicate that there were no artifactual differences caused by the differences in training at the two locations.

Task Performance. The data analyzed in this section focus on task performance only, using an overall task outcome measure and collective task process measures. The primary task outcome measure is the number of rooms successfully completed in a mission scenario, labeled Good Rooms. A successful completion requires that team members search the room, neutralize any opposing forces, check the state of all canisters, and deal with all canisters (disarming any armed canisters) before being called back by the offsite controller due to time constraints. In addition, team members must not have shot any neutral bystanders or exploded any gas canisters. A related collective task process measure is referred to as Search Time, the mean time to search a room (even if errors were made on aspects unrelated to search, in that room).

Repeated measures Multivariate analyses of variance (MANOVAs) were used to address the changes across the missions based on the related measures (Good Rooms and Search Time), and to investigate the differences between local and distributed groups on these identified measures. A significant effect was found for team member location in the MANOVA using Good Rooms and Search Time (Wilks' Lambda, $F\,(2,15) = 5.07$; $p = .021$). A significant effect was also found for the change over missions on these related measures (Wilks' Lambda, $F\,(14,3) = 14.145$; $p = .025$). No significant interaction was found between the distributed nature of the teams and the repeated missions.

The univariate test on Good Rooms only (performed as a part of the MANOVA to test the individual measures), revealed a significant difference over the repeated missions ($F\,(7,112) = 27.264$, $p < .001$). The univariate analysis also revealed a significant difference on Good Rooms between the local and distributed teams ($F\,(1,16) = 10.742$, $p = .005$). This result is also easily discernable in Table 3. No significant interaction was found between the repeated missions and team location. The significant increase in the average number of rooms correctly searched over missions, and the significantly higher scores by the local teams, are shown by the means presented in Table 3.

A similar pattern was found with the average Search Time for rooms. A significant difference was found in the MANOVA over the changes between the first mission and the last ($F\,(7, 112) = 19.787$, $p < .001$) for the teams. The time to search rooms decreased over repeated missions, as shown in Table 4. As with Good Rooms, there was a significant difference between local and distributed teams in their time to search rooms during the missions ($F\,(1, 16) = 6.551$, p

Table 3

Means and Standard Deviations for Number of Good Rooms per Mission by Networking Condition

Mission	1	2	3	4	5	6	7	8	Overall
Local									
M	3.5	5.56	6.78	7.33	6.67	7.78	7.67	8.44	7.5
SD	1.22	1.33	1.72	1.11	1.41	1.86	1.66	1.51	1.12
Distributed									
M	3.33	4.4	5.11	5.89	6.0	6.22	6.33	6.67	6.28
SD	1.0	.88	.78	1.36	1.0	.97	1.32	.71	.62

= .021), with local teams performing faster than distributed teams, as can be seen in Table 4. No significant interaction was found between the repeated missions and team locality on this measure.

Table 4

Means and Standard Deviations for Search Time Per Room by Networking Condition

Mission	1	2	3	4	5	6	7	8
Local								
M	78.78	57.84	51.36	47.31	51.04	43.31	38.87	39.13
SD	22.92	13.62	12.89	12.17	13.74	8.42	8.95	7.31
Distributed								
M	82.15	68.10	63.34	5 7.78	54.75	54.11	50.38	46.39
SD	17.54	19.38	6.13	10.49	9.51	15.44	9.52	4.60

Other collective task process measures were the average time to conduct the collective door entry routine (opening a door, using a concussive grenade, and entering the room, referred to as Door Entry), and the average time to check, disarm, and neutralize armed gas canisters in each mission (a collective task requiring detection of the canister state and code by one member, and disarming the canister by the other, referred to as Canister Disarming). A repeated measures ANOVA was used to investigate the Door Entry routine and a *t*-test was used to analyze the Canister Disarming measure (explained below).

The ANOVA on the average time for Door Entry showed that the times also decreased significantly over repeated missions ($F\ (7, 112) = 10.939, p < .001$), as shown by the means presented in Table 5. Unlike the Search Time and Good Rooms, the decrease in time to perform the Door Entry did not significantly differ between the local and distributed teams. There was no significant interaction found between the repeated missions and team locality on this measure.

Table 5

Means and Standard Deviations for the Time to Open Door and Enter Rooms over Scenarios by Networking Condition

Mission	1	2	3	4	5	6	7	8
Local								
M	12.46	10.7	8.72	8.78	8.55	9.44	8.41	8.19
SD	4.4	3.57	1.47	1.81	1.02	2.27	1.40	.89
Distributed								
M	15.0	8.95	9.42	8.35	8.81	8.57	7.98	8.51
SD	6.72	.92	2.29	1.31	2.66	1.33	.83	1.76

The number of armed canisters encountered during the missions varied, as discussed in the materials and procedures, and therefore the number of canisters that teams disarmed varied by scenario. In addition, the number of teams successfully disarming canisters in the initial missions was low and diverse. (Only five teams successfully disarmed an armed canister in their first mission, and one team did not successfully disarm a canister until the fourth mission.) Therefore an ANOVA on the number of successful canister disarming routines was not appropriate as the number of teams successfully disarming canisters was not equal in every cell. Instead, the average time to disarm the armed canisters was used as the dependent measure for this task in each team's missions. The average time required for teams to successfully disarm discovered (armed) canisters (from checking the canister state through the collective disarming procedure and capping) decreased over mission rehearsals for both local and distributed teams, as shown by the means presented in Table 6. The overall means for the local and distributed teams were calculated across all missions (local = 37.43, distributed = 44.24) and a planned comparison *t*-test was performed, which found no significant difference between the groups ($t(16) = 1.89$, $p = .077$).

Table 6

Means and Standard Deviations for the Time to Disarm Armed Canisters (Detection through Capping) by Networking Condition

Mission	1	2	3	4	5	6	7	8
Local								
M	56.91	46.14	36.72	38.92	39.97	29.44	33.48	30.74
SD	23.91	15.91	11.75	18.55	11.63	7.17	14.47	4.04
Distributed								
M	70.78	59.55	42.87	41.08	45.14	41.69	39.8	35.84
SD	34.32	14.02	11.84	6.64	10.17	10.95	11.36	9.46

The time required to traverse hallways, and the number of collision situations were of interest as indications of improving facility or skill levels within the VE, and indirect indicators of improvement within mission activities. A repeated measures MANOVA found that the

Hallway Movement times decrease significantly over missions ($F = 27.520, p < .001$), as shown by the means in Table 7. There was also a significant difference between local and distributed

Table 7
Hallway Times and Total Collisions by Distributed Condition over Missions

Mission	1	2	3	4	5	6	7	8
Local								
Hallway Times	56.70	42.79	35.26	33.98	38.89	34.40	32.01	29.60
Collisions	9.25	12.56	13.44	12.67	13.44	15.78	16.78	16.89
Distributed								
Hallway Times	68.97	49.38	46.11	41.70	44.61	43.16	42.59	36.94
Collisions	9.22	8.22	8.56	8.78	9.56	10.33	13.56	11.67

team hallway times ($F = 605.99, p < .001$). In addition, a repeated measures MANOVA on the number of collisions made by teams during the missions demonstrated a significant decrease in Collisions over missions ($F= .2.672, p = .047$), but no difference in overall collisions between the local and distributed teams. Means for the Collisions measure are also provided in Table 7.

AAR Communication. Because we were interested in determining if there were differences in communication styles between high and low performing teams while controlling for the differences between local and distributed teams, we performed a median-split on the Good Rooms means for each of the final seven missions for each team type (local and distributed) and dropped the middle performing team for both local and distributed groups. The first mission data was not used in forming the high and low performing groups because this measure was taken before the first AAR, and the analysis was focused on the relationship between AAR communication and subsequent performance. Table 8 gives the Good Room means for the teams in each group.

Table 8
Team Good Room Means over the Final Seven Missions by Performance Group and Networking Condition

	High Performance	Low Performance
Local	8.71, 8.57, 7.43, 7.14	6.29, 6.29, 5.86, 5.86
Distributed	7.14, 6.57, 6.14, 6.00	5.71, 5.57, 5.43, 5.14

In order to determine whether communication patterns were significantly different for locality or performance group, a 2 X 2 MANOVA was conducted. Results showed that team communication patterns across all AARs did not differ significantly between the high and low performance groups for any of the hypothesized communication measures. Further, there were

no significant differences in communications styles between local and distributed teams, nor were there any significant performance group by locality interactions.

As is apparent in data presented in many of the tables, the biggest improvement in performance occurs from mission 1 to mission 2. We therefore hypothesized that examination of the first AAR session (administered between missions 1 and 2) might lend some insight into the mechanisms through which change occurred. Teams were again split in the fashion described above, this time with the difference between number of Good Rooms in mission 1 and in mission 2 used as the dependent variable. This yielded a high improvement local group with an average increase of 3.125 Good Rooms, a low improvement local group with an average increase of 1 Good Room, a high improvement distributed group with an average increase of 2.5 Good Rooms, and a low improvement distributed group with an average decrease of 0.25 Good Rooms. A second 2 X 2 MANOVA showed no main effects in communication patterns of the first AAR for either the improvement group or the locality group. Also, no significant interactions for improvement group by locality group were found.

Team Personality. Prior to testing the five hypotheses concerning performance and personality, we ensured that team members did not differ on personality prior to team tasks. As expected, no significant differences were found, between the five personality factors and a) Location, b) Gender, or c) Team Role. Next, all 18 teams were ranked according to average number of Good Rooms over the eight missions. From this ranking, we grouped the top four teams as "high-performing" and the bottom four as "low-performing." Therefore, eight teams were included in the analyses below.

A one-way MANOVA was then used to examine the personality factors measured by the NEO-FFI, with team performance serving as the independent variable (high-performing vs. low-performing) and team average scores on the five personality factors serving as the dependent variables. Results showed a significant main effect for team performance on extraversion, (F (1, 6) = 13.15, p = .011, η^2 = .69). Mean team personality scores for high- and low-performing teams are shown in Table 9.

Table 9
Mean Team Personality Scores for High- and Low-Performing Teams

	High-Performing			Low-Performing		
Factor	*M*	*SD*	*n*	*M*	*SD*	*n*
Neuroticism	47.38	4.61	4	49.75	3.93	4
Extraversion	58.38	2.50	4	48.88	4.61	4
Openness	62.63	3.45	4	53.50	8.35	4
Agreeableness	45.13	4.55	4	40.00	4.24	4
Conscientiousness	49.63	6.51	4	44.13	10.49	4

A second one-way MANOVA was performed to test the hypothesis regarding the teams' personality diversity, or TPD, on the extraversion and neuroticism personality factors. As with the first MANOVA, team performance served as the independent variable. For this analysis, however, mean team personality diversity for each of the five personality factors served as the dependent variables (DV). Each DV was calculated by averaging difference scores — representing the difference between a team leader's score on a each personality factor and the equipment specialist's score on the same factor — for both high- and low-performing teams. Results did not support the hypothesis that high TPD on extraversion and neuroticism would be associated with better team performance. Mean team personality diversity scores for high- and low-performing teams are shown in Table 10.

Table 10
Mean Team Personality Diversity Scores for Personality Factors by Performance Grouping

	High-Performing			Low-Performing		
Factor	*M*	*SD*	*n*	*M*	*SD*	*n*
Neuroticism	9.75	6.85	4	8.00	8.00	4
Extraversion	8.25	7.41	4	12.75	6.55	4
Openness	12.25	9.61	4	4.50	7.72	4
Agreeableness	15.25	10.24	4	19.00	7.02	4
Conscientiousness	18.25	13.50	4	8.75	4.27	4

Discussion

The major focus of this research was the investigation of changes in task and team outcome and process measures over mission rehearsal trials, and whether the measures revealed any differences between local and distributed teams. The expectation was that all teams would improve over the course of repeated mission rehearsals. The interesting issue was whether the distributed teams would perform differently overall, or would demonstrate a different pattern of improvement during the repeated mission rehearsals. These performance differences, if any, would be based on differences in the team interactions that would arise from the nature of the distributed team situation. Accordingly, the discussion of analysis outcomes will begin with the mission rehearsal performance results, then review the findings from the ancillary questionnaires and measures, and finish by discussing the findings from the training sessions and the relationship of training to the mission rehearsal data.

Mission Rehearsals

Task Performance. The discussion in this section focuses exclusively on the mission rehearsal tasks, not team characteristics (e.g., communication, personality make-up) or reaction

to the VE (e.g., from the training sessions). The task information concerns the dependent variables for successful completion of mission steps or tasks (e.g., Good Rooms and Canister Disarming), and the overt performance process measures (Search Time, Hallway Movement, and Door Entry). The results provide clear evidence that task performance improves over mission trials. This finding, that teams improve in both task outcome and task process measures over repeated trials, is not surprising. Humans that practice nearly any task, with attention toward improvement and feedback about performance, will improve their performance.

More importantly, the results also show significant differences between local and distributed teams on several measures. Local teams performed better in terms of Good Rooms, Search Time, and Hallway Movement times. However, the Door Entry, Collisions, and Canister Disarming routines were not significantly different between the experimental groups, with Door Entry improving over missions while Collisions became more frequent. These performance results are discussed in reverse order from the analysis presentation, as the reasoning for each of the outcomes aids and supports the overall rationale presented for all of the outcomes.

Collisions, unlike the other measures, increased significantly over the repeated missions. During the initial movement training, collisions were treated and counted as errors. When collisions occurred in training, participants were coached in recovery techniques and told to try to decrease the number of collisions as they would slow the participant during the mission rehearsal exercises. Once the team mission rehearsals began, collisions were essentially ignored, unless a participant was making extreme errors that were seriously delaying the team. Based on the patterns of the different data collected, the problem is why the significantly improving overall outcome measure of successful room searches and the related search time measure, as well as hallway movement times, were not sensitive to the increasing number of collisions. The only tenable explanation is that the collision states became minimally intrusive and perhaps even useful in guiding performance. In support of this explanation, we noted that as teams became more efficient some participants began to collide with the hallway wall as a technique to ensure that they were in proper position for executing the door entry routine.

In retrospect, this should have been expected as it fits our experience over the course of developing and testing different VE configurations and graphics. We have witnessed that recovery from collision states seems to become easier the longer a participant is in a VE. In the present study, participants actually collided more frequently over time, even though processes in the VE were not negatively affected. For example, collisions did increase time-based activities such as movement through hallways. We therefore argue that participants learned to use wall and object collisions as another source of information about the VE, similar to the way in which real life physical contact with objects is often used as an additional feedback option for performance.

Finding no differences between local and distributed teams on the Door Entry and Canister Disarming measures may be explained by examining the nature of these two tasks. These tasks require close interaction between the team members within a rigid task format. In each of these collective tasks, each team member would get ready to perform and communicate

their readiness (before our overt performance measurement could begin); one would begin the activity, and the other would closely follow with their portion of the task, with alternating actions continuing until the task was completed. Therefore, any mistakes in performance or timing could easily be identified immediately during the activity, or pointed out during the AAR after the mission, and relatively easily rectified in future performance in either case. There would not necessarily be any need for external feedback, for example during the AAR, from one's team member or an examination of the protocol sheet (given as a guide during AARs), in order to improve in the task. In fact, there may have been a reluctance to acknowledge or extensively discuss errors. The ease with which each team member could identify and correct his or her own portion of the tasks would tend to decrease any differences between local and distributed team performance. This factor might also serve to minimize any discussion about conditions or error states, as noted below in the discussion of the communication results.

The argument presented for both groups uniform improvement with Door Entry, and the finding of no significant difference between team distributions in Canister Disarming, also provides a framework for explaining the significant difference in improvement between local and distributed teams on the Good Rooms, Search Time, and Hallway Movement measures. These three measures address combinations of activities, in some of which the errors are not obvious or easy to monitor, either by the participant or the participant's team member. The Room Search and Hallway Movement activities require flexible and coordinated movement while searching or covering an area, and possibly identifying and dealing with opposing forces. Each of the activities can certainly improve somewhat through self-monitoring, but the inclusion of less structured activities in the collective set probably requires a higher level of monitoring and more team coordination. This higher level of monitoring, feedback, and planned team correction and coordination might be easier to effect or initiate in the local condition.

Since the only apparent difference between the teams was the capability for face-to-face, <u>between-mission</u> interaction, we conclude that something associated with the face-to-face interaction supported superior performance on certain types of tasks by the local team. This difference in performance seemed to arise after the first mission (see the means in Tables 3, 4, and 7) and was not erased by further feedback and practice opportunities. Based on the performance measures, we cannot be certain what is behind this difference, however, it is likely that communication patterns of the distributed teams are partly responsible. Both in-mission and AAR communication data were collected and may provide insight into these findings. The within-mission communication has yet to be analyzed, the AAR communication data are discussed next.

AAR Communication. Finding no significant differences in communication styles between local and distributed teams or high and low performing teams in this particular experiment indicates that the difference in performance between local and distributed teams may not be due to differences in any of the AAR communication patterns examined so far. Distributed teams were limited in that they could not communicate via face-to-face interaction during the AAR, however specific patterns of content and response analyzed to date do not differ.

It is possible that the means by which the communication data were captured or aspects of the coding scheme may not have been sensitive to the relationship between communication style and locality. One possible explanation for the finding that local and distributed teams do not differ in the degree to which they close loops (which counted total communications, planning statements, or mission-related questions) is that the face-to-face interaction gave local teams the opportunity to respond nonverbally, while the distributed teams were not afforded this opportunity. We captured the communication data used for analyses on audio tape only, thus were absent of any nonverbal communications, such as head shaking.

If the local teams could communicate visually, then some number (perhaps a significant number) of closed communications actually occurred and supported or led to the improved performance. Thus communications that were truly closed would have been coded as open (single utterances) for these teams. Also, in this analysis of communication, laughter was not coded as a response, although in many instances it may have served this purpose. It is possible that, if a relationship does indeed exist between communication and performance, missing these and other critical pieces of information for certain teams might have weakened the validity of our communication measures and the lessened the possibility of determining significant factors behind the team differences.

It is interesting that although Bowers et al. (1998) found that the in-flight communications patterns of good teams differed from those of poor-performing teams, the results were not replicated here for measures of between-mission AAR communications. One reason for this may be that using certain types of communication during the more structured format of an actual mission with time constraints is more important than using these communication patterns between missions, when teammates can devote more attention to each other and the task. During a 10-minute between-mission interaction period, a team is not under these crucial time restraints and communications may not need to be so structured and efficient to optimize performance. There is plenty of time, for example, for teammates to ask for and gain clarification. It may not be so important that teammates respond to all questions, commands, etc., immediately in a between-task interaction as there is ample time for the communication attempt to be repeated. Future research might examine the relationship between performance and communication patterns during a much shorter between-task segment. It can also be again noted that losing laughter—which was not scored in the verbal transcriptions—and nonverbal communications as responses weakens the sensitivity of our measures and reduces the possibility of finding causal effects.

Another interesting finding is that we did not find effective team planning to be associated with better team performance, as others have found previously (Orasanu, 1990; Stout et al., 1999). One possible explanation is that in this research planning behavior was measured in terms of planning communication counts, proportion of conversation devoted to planning, and proportion of planning utterances that were responded to. Stout et al. measured the quality of planning in nine planning dimensions rather than communication patterns. It is possible that planning would have been predictive of performance if measured in terms of type or quality,

rather than quantity of overall planning communications. Time and effort have precluded that analysis as of this writing.

Finally, it is possible that communication styles do differ between local and distributed teams, or between high and low performing teams, on a certain number of characteristics that we did not examine in the present experiment. For example, future research might look at the number of supportive statements given by teammates, leadership exhibited in communication, or the quality and accuracy of information transmitted.

Neither the hypothesized differences between local and distributed teams, nor the hypothesized differences between high and low performing teams were obtained. This indicates that locality does not influence between-mission communication style (for those variables measured in the present study), nor does between mission communication influence performance. It also suggests that the difference in performance between local and distributed teams is not a function of communication during the AAR.

It is clear that more research is needed to determine the nature of the relationship between communication patterns, locality, and performance. Perhaps the analyses of in-mission communication patterns will reveal differences between the communication styles of local and distributed teams. It may be the case that although locality does not influence between-mission communication, it has an affect on within-mission communication. For example, teammates who have not met or communicated face-to-face may not be as comfortable with each other during a distributed mission session and may be hesitant to correct teammate's mistakes, to provide feedback, or to use instructive language. Researchers might also want to examine the specific effects caused by a loss of nonverbal communication in a distributed interaction situation. Much more exploration of the role of communication in team performance will be necessary to help us understand this complex variable.

Team Personality. With regard to personality, only the hypotheses addressing Extraversion, that high performing teams will exhibit higher mean levels of Extraversion than low-performing teams, was supported by the data. No significant differences were found between low- and high-performing teams on average levels of Conscientiousness, Neuroticism, or Agreeableness. Furthermore, there were no differences in the diversity of team members' scores on Extraversion and Neuroticism.

The significant Extraversion finding supports the arguments of others (e.g., Costa & McRae, 1992) that Extraversion plays a role in interpersonal interactions. Because the team tasks required cooperation, communication, and team interaction, it follows that teams with members that are more prone to initiating and continuing interpersonal interactions, as suggested by high Extraversion scores, would perform at a higher level. It is unclear, however, why no significant difference was found on the Agreeableness factor, the other personality factor attributed to success in interpersonal interactions. Although mean Agreeableness scores were higher for good teams, we expected a much larger, and significant, difference. Additional

research examining Extraversion and Agreeableness in team performance is needed to determine the role these traits have in team interpersonal interactions and cooperation.

Beyond the Extraversion finding, the present results did not completely agree with extant literature on team performance and personality. One possible reason for this is the size of the teams used in our study. Teams comprised of two individuals may interact and behave differently than larger teams. For instance, in the aforementioned Neuman et al. (1999) study, which found that high team personality diversity on Extraversion and Neuroticism was related to better performance, the experiment involved four-person teams.

Another potential source of the difference between the present findings and previous research is a time factor. Most studies on personality only assess participants for short time periods (Chidester et al., 1991). As a result, performance in the short-term might be differentially sensitive to personality effects when compared to performance after longer exposure to operational or experimental conditions. For example, Helmreich, Sawin, and Carsrud (1986) studied airline pilot performance immediately after training and after 6 months on the job. Personality measures did not predict performance evaluations made after training, but after 6 months, personality became significantly correlated with performance, a finding the authors termed a "honeymoon effect." They proposed that people are initially motivated to do as well as possible when starting a new job or task. Later, however, as the task becomes routine, initial motivation declines and personality characteristics may surface that affect performance.

Based on these findings, a defensible position is that personality-performance relationships found in short-term studies cannot be compared to findings from longer-term studies. In the present study, participants were evaluated over a relatively long time period (2, 4-hr sessions spread over 2 days), longer than in many personality-performance investigations which often take a "snapshot" of performance and relate this brief assessment to personality. Accordingly, time may be partially responsible for the fact that our findings were dissimilar from earlier research on team performance and personality.

Results of the present study suggest further research into team performance and personality is warranted. Plausible approaches include forming teams based on member personality attributes, manipulating team size, extending the performance observation period, and assessing team performance in other domains.

VE Training

In order to ensure that any discovered differences were based in the distributed nature of the simulation, several VE training sessions were provided for all participants on all aspects of the experimental scenarios. In addition, standard biographical information, and questionnaires that address people's reactions to VE technology was collected before and during both training and mission rehearsal sessions.

As described in the introduction, training was standardized for all participants. The major measured variable for the training was the number of VE sessions used to train participants to criterion in VE and task-specific operations. As shown in the results, the participants trained at SSRU differed slightly, but significantly, from the smaller number (10) trained at DCIEM. When the participants were combined into teams, with random assignment of roles on the teams, there was no significant difference between the local teams and the distributed teams. Because the performance data were analyzed by team only, this result would seem to support the validity of our mission rehearsal results.

Conclusions

The research reported here required considerable resources and expertise in order to provide the necessary and sufficient conditions for addressing possible training and performance differences between local and geographically separated teams. The experiment required non-standard, exploratory technology in order to address issues that would not have been feasible to address in any other way. This technology will become the standard, and even become commercial off-the-shelf (COTS), by the end of the decade. The time for empirically based recommendations about the use of technology is not after it becomes commercially available. The most important time for training effectiveness recommendations is during exploratory development and initial fielding. Yet without developing cutting edge technology for exploring learning and training effects, there can be no empirically based recommendations, only guesses. On that basis, this experiment is a significant accomplishment. It shows not only that geographically distributed virtual simulations for individuals are feasible, but also that the technology can be the basis for psychological investigations.

It is clear that geographically distributed training and rehearsal in VE simulations will be developed and implemented for training in the future. The central issue is whether teams comprised of geographically distributed individuals learn and improve in the same manner and amount as teams whose members can all interact locally. Our research shows that even though all participants are learning and rehearsing the same missions in the same simulations, there are subtle intervening variables that can lead to less effective training when geographically separated teams are involved. These issues are critical to the development and fielding of extensively distributed systems for training dismounted soldiers. Our data show significant differences in both outcome and process task measures. This information should be used to help find ways to diminish or eliminate any possible differences in the performance of locally comprised versus extensively distributed teams. At the risk of employing an old cliche, more research is needed, and requires the use of leading edge technology and development in order to answer seemingly simple questions. A possible next step is to revisit the distributed team settings. For example, manipulating the dimensions for team interpersonal interactions, by adding video to the distributed teleconferencing, might provide a clearer picture of how this interpersonal interactions affects team performance. Another interesting and more practical approach would be to examine the benefits of a brief team interaction training session. Local and distributed teams could be given instruction on team monitoring and communication skills, and compared with control groups without team communication training. This would determine whether

simple training solutions can "balance out" the local/distributed differences obtained in the present study.

The final and most critical point is that training and mission rehearsals must be equivalently effective, even when some teams are local and others are distributed. When distributed virtual simulation is used for complex collective and cooperative tasks, the system or instruction must counter the differences we have found between local and distributed teams. The next step should be to find ways to alleviate or counter these discovered differences. The training analysis and development process can be eased if the developer knows how to alter the instructional approach to include or emphasize those team cohesion and communication skills necessary to alleviate any differences arising from the distributed simulation.

References

Aiken, L. R. (1989). *Assessment of personality.* Boston: Allyn and Bacon.

Atwater, L. E. (1992). Beyond cognitive ability: Improving the prediction of performance. *Journal of Business and Psychology, 7,* 27-44.

Atwater, L. E., & Yammarino, F. J. (1993). Personal attributes as predictors of superiors' and subordinates' perceptions of military and academic leadership. *Human Relations,* 46, 645-668.

Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology, 83*(3), 377-391.

Bing, M. N., & Lounsbury, J. W. (2000). Openness and job performance in U.S.-based Japanese manufacturing companies. *Journal of Business & Psychology, 14*(3), 515-522.

Blake, R. J., Potter, E. H., III, & Slimak, R. E. (1993). Validation of the structural scales of the CPI for predicting the performance of junior officers in the U. S. Coast Guard. *Journal of Business and Psychology, 7,* 431-448.

Bowers, C., Jentsch, F., Salas, E., & Braun, C. (1998). Analyzing communications sequences for team training needs assessment. *Human Factors, 40,* 672-679.

Brown, B. R., Nordyke, J. W., Gerlock, D. L., Begley, I. J., & Meliza, L. L. (1998). *Training Analysis and Feedback aids (TAAF Aids) study for live training support* (Study Report 98-04, ADA351107). Alexandria, VA: United States Army Research Institute for the Behavioral and Social Sciences.

Chidester, T. R., & Foushee, H. C. (1989). Leader personality and crew effectiveness: A full-mission simulation experiment. In R. S. Jensen (Ed.), *Proceedings of the Fifth International Symposium on Aviation Psychology: Vol. II,* 676-681.

Chidester, T. R., Helmreich, R. L., Gregorich, S. E., & Geis, C. E. (1991). Pilot personality and crew coordination: Implications for training and selection. *International Journal of Aviation Psychology, 1,* 25-44.

Comrey, A. L. (1953). Group performance in a manual dexterity task. *Journal of Applied Psychology, 37,* 207-210.

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO-PI-R professional manual: Revised NEO personality inventory (NEO PI-R) and NEO five-factor Inventory (NEO-FFI).* Odessa, FL: Psychological Assessment Resources.

Driskell, J. E., & Salas, E. (1992). Collective behavior and team performance. *Human Factors, 34*, 277-288.

Driskell, J. E., Hogan, R., & Salas, E. (1988). Personality and group performance. *Review of Personality and Social Psychology, 14*, 91-112.

French, E. (1958). Development of a measure of complex motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society* (pp. 242-248). Princeton, NJ: Van Nostrand.

Gersick, C. J. (1988). Time and transition in work teams: Toward a new model of group development. *Academy of Management Journal, 31*(1), 9-41

Goldstein, I. L. (1974). *Training: Program development and evaluation.* Monterrey, CA: Brooks/Cole Publishing Co.

Hackman, J. R. (1993). Teams, leaders, and organizations: New directions for crew-oriented flight training. In E. L. Wiener, & B. G. Kanki (Eds.) *Cockpit resource management.* Academic Press, Inc, San Diego, CA, p. 47-69.

Helmreich, R. L., Sawin, L. L., & Carsrud, A. L. (1986). The honeymoon effect in job performance: Temporal increases in the predictive power of achievement motivation. *Journal of Applied Psychology, 71*, 185-188.

Jacobs, R. S., Crooks, W. H., Crooks, J. R., Colburn, E., Fraser, II, R. E., Gorman, P. F., Madden, J. L., Furness, III, T. A., & Tice, S. E., (1994, October). *Behavioral requirements for training and rehearsal in virtual environments* (TR-1011). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Jackson, S. E. (1992). Team composition in organizational settings: Issues in managing an increasingly diverse workforce. In S. Worchel, W. Wood, & J. Simpson (Eds.), *Group processes and productivity* (pp. 138-173). Newbury Park, CA: Sage.

Jentsch, F. G., Sellin-Wolters, S., Bowers, C. A., & Salas, E. (1995). Crew coordination behaviors as predictors of problem detection and decision making times. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting* (pp. 1350-1353). Santa Monica, CA: Human Factors and Ergonomics Society.

Johnston, J. H., Smith-Jentsch, K. A., & Cannon-Bowers, J. A. (1997). Performance measurement tools for enhancing team decision-making training. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp.311-330). Mahwah, NJ: Lawrence Erlbaum Associates.

Kanki, B. G., Lozito, S. C., & Foushee, H. C. (1989). Communication indices of crew coordination. *Aviation, Space, and Environmental Medicine*, *60*, 56-60.

Kickul, J., & Neuman, G. (2000). Emergent leadership behaviors: The function of personality and cognitive ability in determining teamwork performance and KSAS. *Journal of Business & Psychology, 15*(1), 27-51.

Knerr, B. W., Goldberg, S. L., Lampton, D. R., Witmer, B. G., Bliss, J. P., Moshell, F. M. & Blau, B. S. (1994). Research in the use of virtual environment technology to train dismounted soldiers. *Journal of Interactive Instruction Development, 6*(4), 9-20.

Levison, W. H., & Pew, R. W. (1993, February). *Use of virtual environment training technology for individual combat simulation* (ARI-TR-971, ADA263546). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43*, 335-354.

Meliza, L. L., Bessemer, D. W., & Hiller, J. H. (1994). Providing unit training feedback in the distributed interactive simulation environment. In R. F. Holz J. H. Hiller, & H. H. McFann, (Eds.), *Determinants of effective unit performance: Research on measuring and managing unit training readiness* (pp. 257-280). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences.

Moreland, R. L., & Levine, J. M. (1992). The composition of small groups. *Advances in Group Processes, 9*, 237-280.

Morgan, B. B., Jr., Salas, E., & Glickman, A. S. (1993). An analysis of team evolution and maturation. *The Journal of General Psychology, 120*, 277-291.

Mosier, K. L., & Chidester, T. R. (1991). Situation assessment and situation awareness in a team setting. In Y. Queinnec & F. Daniellou (Eds.), *Designing for everyone: Proceedings of the 11th Congress of the International Ergonomic Association* (pp. 798-800). London: Taylor & Francis.

Neuman, G. A., & Wright, J. (1999). Team effectiveness: Beyond skills and cognitive ability. *Journal of Applied Psychology, 84*(3), 376-389.

Neuman, G. A., Wagner, S. H., & Christiansen, N. D. (1999). The relationship between work-team personality composition and the job performance of teams. *Group and Organization Management, 24*, 28-45.

Orasanu, J. M. (1990). *Shared mental models and crew decision-making* (CSL Tech. Report No. 46) Princeton, NJ: Princeton University, Cognitive Science Laboratory.

Reeves, G. (2001). Street smarts. *Military Training & Technology, (6)*3, 18-21.

Rosse, J. G., Miller, H. E., & Barnes, L. K. (1991). Combining personality and cognitive ability predictors for hiring service-oriented employees. *Journal of Business & Psychology, 5*, 431-445.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. L. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3-29). Norwood, NJ: Ablex.

Schneider, F. W., & Delaney, J. G. (1972). Effect of individual achievement motivation on group-problem-solving efficiency. *Journal of Social Psychology, 86*, 291-298.

Stout, R. J., Cannon-Bowers, J. A., Salas, E., & Milanovich, D. M. (1999). Planning, shared mental models, and coordinated performance: An experimental link is established. *Human Factors, 41*(1), 61-71.

Stout, R. J., Salas, E., & Carson, R. (1994). Individual task proficiency and team process behavior: What's important for team functioning? *Military Psychology, 6*(3), 177-192.

Terborg, J. R., Castore, C., & DeNinno, J. A. (1976). A longitudinal field investigation of the impact of group composition on group performance and cohesion. *Journal of Personality and Social Psychology, 34*, 782-790.

Tesluk, P., Mathieu, J. E., Zaccaro, S. J., & Marks, M. (1997). Task and aggregation issues in the analysis and assessment of team performance. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, methods, and applications* (pp. 197-224). Mahwah, NJ: Lawrence Erlbaum Associates.

United States Army: Simulation, Training, and Instrumentation Command (STRICOM) (n.d.). *Close Combat Tactical Trainer (CCTT)*. Retrieved August 7, 2001 from http://www.stricom.army.mil/STRICOM/PM-CATT/CCTT/.

Vickers, R. R. (1995). Using personality assessment for leadership selection. *U.S. Naval Health Research Report* (No. 95-16), 1-35.

Appendix A. Simulator Sickness in Virtual Environments

Simulator Sickness

Research conducted in the ARI Simulator Systems Research Unit VE program and elsewhere (e.g., Lampton et al., 1994; Wann, 1993) has indicated that simulator sickness is often associated with exposure to VE systems. Simulator sickness is a fairly common phenomenon in which participants suffer one or more physical symptoms (slight through severe) as a direct result of exposure to the simulator, either during or after the interaction. It is possible that the occurrence of these simulator sickness symptoms may affect research outcomes. On this basis, simulator sickness is one issue that we regularly investigate in experiments within our program. We use the self-report SSQ developed by Kennedy et al. (1993). These researchers used a factor analysis of scores on many symptoms collected in different situations to identify three subscales of simulator sickness symptoms, and a combined total severity scale. The scales are all derived by summing the severity scores for a set of symptoms and weighting those sums (using a different weight for each scale). The *Nausea* scale symptoms are drawn from the ratings on general discomfort, increased salivation, sweating, nausea, difficulty concentrating, stomach awareness, and burping. The *Oculomotor Discomfort* scale reflects the ratings from problems with general discomfort, fatigue, headaches, eyestrain, difficulty focusing, difficulty concentrating, and blurred vision. The *Disorientation* scale addresses the ratings on difficulty focusing, nausea, fullness of head, blurred vision, dizziness with eyes open, dizziness with eyes closed, and vertigo. The *Total Severity* score is a sum of the subscale symptom sums weighted with a separate value. Over the course of several experiments, we have used these scales to measure the sickness caused by our VE systems, and endeavored to decrease or minimize the simulator sickness symptomology of experimental participants through control of experimental procedures (e.g. Singer, Ehrlich, & Allen, 1998).

In the earliest experiments conducted in our program, we found simulator sickness to be linked to time in the VE (Knerr et al., 1994). Therefore in the later experiments we have limited the amount of time people spend wearing head-mounted displays (HMDs) and performing experimental tasks (as also recommended by McCauley & Sharkey, 1992). Over the course of repeated short-duration trials, we noted that the greatest change in symptoms occurs early in the experiment rather than later (Singer, Ehrlich, & Allen, 1998), which seems to follow the pattern with sickness in simulators. Further, there is evidence that simulator sickness is lessened as experience with the simulator increases (McCauley & Sharkey, 1992; Lampton, Kraemer, Kolasinski, & Knerr, 1995). For example, Lampton et al. (1995) studied simulator sickness in a tank driver trainer under non-experimental conditions (non-interference in the training program, without selection of students or manipulation of training conditions). The tank trainer, used for initial driver training, had a visual display and a six-degree of freedom motion platform. Lampton et al. used the SSQ scales to measure symptom levels before and after training sessions using the tank driver trainer. The analyses reported significant post-exposure SSQ score differences for the Total, Nausea, and Disorientation scales between the first two sessions on the trainer and the mid-course or the last session scores. The observed decrease in the SSQ scores

over the course of training experiences indicates that the students were adapting to the tank driver trainer (Lampton et al., 1995).

A review by Kolasinski (1995) identified individual, equipment, and task variables that can influence the incidence of simulator sickness. Kolasinski concluded that the results of the reviewed research provides a good basis for hypotheses about sickness that occurs in VE, and that the practicalities of VE research mean that research on simulator sickness in VE will be ancillary. Kolasinski identified a wide range of factors associated with simulator sickness and classified them into three major categories: individual, task, or equipment (simulator) based. Another consideration, as McCauley and Sharkey (1992) point out, is that much of the research on simulator sickness has been conducted on a self-selected and screened sample of the normal human population, pilots. Pilots in the armed services are motivated and have been trained to adapt to extreme motion, with less adaptive individuals not meeting basic criteria and "washing out." The normal population of VE users will presumably go through the same selection process, although it will probably be a less restrictive process. In the interim, individual factors such as age, gender, mental abilities, and other personal characteristics need to be investigated (Kolasinski, 1995). General task characteristics such as degree of control, duration of experience, global visual flow, and head movements (Kolasinski, 1995), can also affect simulator sickness severity. This task characterization emphasizes the need for investigation of simulator sickness across many task domains. Perhaps the most important category of simulator sickness factors is VE equipment characteristics, which include position-tracking error, visual display characteristics, scene content, etc. (Kolasinski, 1995).

It was not the intent of this research to directly manipulate equipment or task variables in an attempt to identify their contribution to VE sickness. We did not anticipate differences in simulator sickness to interact with the distributed nature of the experiment. However, the administration of multiple short VE sessions (see Methods section in the body of this report) provided the opportunity to administer the SSQ repeatedly during the experiment. In particular, it provided the opportunity to address the onset and course of symptoms (during our training phase), which has previously been shown to increase most rapidly during initial sessions and plateau or reduce over subsequent exposures (Singer, Ehrlich, & Allen, 1998). The earlier research suggested that participants adapt to VE with lower levels of induced sickness, although none of the identified major parameters were manipulated (Kolasinski, 1995). Based on earlier research findings, we hypothesized that there would be a significant increase in symptomology over the initial VE session and that the symptoms would reduce to near normal after a 30-minute recovery period after the initial session. A further expectation was that as participants adapted to the VE configuration and task requirements, their change in symptom level over multiple VE sessions would diminish with repeated exposures.

Methods

Participants

As described in the body of the report, participants were acquired at two geographical locations; Orlando, FL, USA, and Toronto, Canada. There were sixty-four training participants with a median age of 21. There were thirty-six participants assigned to teams for the mission rehearsal phase of the experiment, with a median age of 21.

Materials

As noted in the body of the report, all questionnaires were administered via computer using an Access[tm] database. The SSQ questionnaire is replicated in Appendix F.

Procedures

The SSQ was administered before and after every VE session throughout training and the team mission rehearsals. The questionnaire was also administered thirty minutes after the last VE exposure at the end of each session. This insured that no participant left the experiment with elevated symptom levels. On the rare occasion that a participant experienced dramatically elevated levels, they were kept on-site until their levels diminished to near normal, within levels based on the norms provided in Kennedy et al., 1993). These individuals would repeatedly complete the SSQ (approximately every thirty minutes) until their scores were acceptable.

Results

Each SSQ symptom (Kennedy et al., 1993) is scored zero to three for symptom levels none to severe, respectively. Even though multiple symptoms are summed and weighted to form separate scales (Nausea, OculoMotor Discomfort, Disorientation, and Total Severity), the scales are often at zero because the participant does not report any symptoms. As a result, the SSQ data for the groups are not normally distributed. A presentation of summary information about all administrations of the SSQ is not informative due to the large number of individual SSQs (some participants filled out the questionnaire over 30 times during the course of the entire experiment). As discussed in the introduction, certain comparisons are of interest, primarily the changes associated with the initial VE exposures. Descriptive statistics for the training VE sessions are provided in Table 1.

Wilcoxon Signed Ranks Tests were conducted on the change in SSQ scores associated with the first VE exposure (pre versus post). A significant increase from pre-VE to post-VE was found in the Nausea subscale ($Z = -2.303$, $p = .021$), and the Disorientation subscale ($Z = -3.229$, $p = .001$). The Wilcoxon test was also used to examine the differences between the first and second VE exposures, by comparing the amount of change (signed differences, pre minus post) over the exposures on the scales. This comparison found only the change in the Disorientation

scale being significantly different ($Z = -2.281, p = .023$), indicating that there was significantly less change in the Disorientation scale over the second VE exposure than the symptom change that resulted from the first VE exposure.

Table 1
Means and Standard Deviations for SSQ Score Statistics for the First, Second, and Last VE Exposures during Training

	1st VE Session		2nd VE Session		Last VE Session	
	Pre-	Post-	Pre-	Post-	Pre-	Post-
Total Severity						
M	7.70	11.95	6.65	6.56	5.42	6.25
SD	10.0	15.08	10.24	10.64	9.68	9.08
Nausea						
M	5.84	11.68	6.21	5.75	3.56	5.55
SD	9.68	19.14	11.89	12.06	9.37	10.12
Disorientation						
M	3.74	11.84	3.98	7.07	4.57	5.19
SD	8.93	17.52	9.16	14.33	11.97	9.64
OculoMotor Discomfort						
M	9.94	8.60	6.38	4.93	5.66	5.43
SD	11.27	10.13	8.84	8.19	8.79	8.00

Mission Simulator Sickness. Obviously only those trainees that adjusted to the VE completed training, and as a result there were no dropouts during the mission sessions. The overall SSQ scores did not vary dramatically during the repeated VE sessions. The change in SSQ scores from before to after the first mission was limited, with only 6 out of 36 participants changing over the course of the first mission. The change over the last (eighth) mission was also minimal, with only 9 out of 34 (2 questionnaire sets missing) changing in any way. No analyses were conducted on these changes because the largest proportion of the subjects did not record any change over missions.

Discussion

The Wilcoxon Signed Ranks Test was used to examine the changes in SSQ scales in response to the multiple VE exposures during training, using the ranked and signed differences between pre and post VE measures. When a significant difference is found using this test, it indicates that the matched groups do not have the same distributions. The conclusions cannot be

drawn about the means of the two groups, as the distributions are not normal (Hays, 1973). Our conclusions must be (conservatively) drawn about the entire distribution of observed scores or differences. Since there are a large number of ties or zero changes in the groups, we must look also at the actual changes in order to draw any conclusions. For the change between the pre and post scores associated with the initial VE exposure, the observed trends in the SSQ indicate that while a few people improved (decreased their symptom reports), a larger proportion became more symptomatic (increased their reported symptom levels). This is taken to mean that the initial exposure creates increased distress in a substantial proportion of the population, which matches the results and interpretations from all of the research literature on both simulators and VE systems. The decrease in the Disorientation scale also supports much of the research literature, although the consistently elevated Nausea scale findings suggests that the adaptation is slower for the symptoms measured by that scale.

As noted in the results, above, only those participants that could finish the training successfully could be assigned to a team. Therefore, there was no reason to expect that simulator sickness symptoms would either increase or decrease over the repeated missions. The minimal proportion of SSQ scores that changed pre-to-post over the first (6 out of 36) and last mission (9 out of 34) would seem to support these hypotheses. Moreover, the large number of constant responses, pre and post exposure, leaves data that is nearly impossible to analyze. The most obvious non-parametric test, the Wilcoxon Test, which uses the signed ranks (Hays, 1973), is not appropriate because the number of tied ranks (no changes) distorts the interpretation of results (see Hays, 1973 for a short discussion).

An inspection of the responses to the SSQ during training and over the repeated missions indicates that our regime (12 minutes in the VE with a 30 minute recovery period) is a successful one. Success in this case means that a large proportion of the population adapts to the VE, and are not troubled by further repeated sessions. This is excellent information for researchers that are interested in using VE in situations which can fit the restricted time segments. Situations that require longer periods in a VE, or shorter recovery times, will probably still run afoul of the increased symptoms typical with longer term or more frequent use in simulators. It is not clear what will happen in this domain as the VE equipment improves to support significant gains in realism.

Appendix A: References

Hays, W. L. (1973). *Statistics for the social sciences.* New York: Holt, Rinehart, & Winston.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). A simulator sickness questionnaire (SSQ): A new method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*(3), 203-220.

Knerr, B. W., Goldberg, S. L., Lampton, D. R., Witmer, B. G., Bliss, J. P., Moshell, F. M. & Blau, B. S. (1994). Research in the use of virtual environment technology to train dismounted soldiers. *Journal of Interactive Instruction Development*, *6*(4), 9-20.

Kolasinski, E. M. (1995). *Simulator sickness in virtual environments* (ARI-TR-1027, ADA 295 861). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Lampton, D. R., Kolasinski, E. M., Knerr, B. W., Bliss, J. P., Bailey, J. H., & Witmer, B. G., (1994). Side effects and aftereffects of immersion in virtual environments. *Proceedings of the Human Factors Society 38th Annual Meeting* (pp. 1154-1157). Santa Monica, CA: Human Factors and Ergonomics Society.

Lampton, D. R., Kraemer, R. E., Kolasinski, E. M., & Knerr, B. W. (1995). *An investigation of simulator sickness in a tank driver trainer* (ARI Research Report 1684, ADA 304 277). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

McCauley, M. E., & Sharkey, T. J., (1992, Summer). Cybersickness: Perception of self-motion in virtual environments. *Presence, 1*(3), 311-318.

Singer, M. J., Ehrlich, J., Allen, R. C. (1998). *Effect of a body model on performance in a virtual environment search task.* (Technical Report 1087, ADA 352 026). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Wann, J. (1993). What's wrong with your head mounted display? *CyberEdge Journal, 5*(Supplement 17).

Appendix B. Presence and Immersion

Presence and Immersion

The efficacy of VEs has often been linked in the literature to the sense of presence experienced in those VEs, although there are arguments about how to measure presence, and insufficient evidence to show that it directly affects performance (see Witmer & Singer, 1998). Presence is defined as the subjective experience of being in one place or environment, even when one is physically situated in another (Witmer & Singer, 1994, 1998). Witmer and Singer have developed and refined subjective questionnaires that address a person's baseline immersive tendencies (the Immersive Tendencies Questionnaire, ITQ) and people's responses to the VE situations used in the research program (the Presence Questionnaire, PQ).

Although the concept of presence has been widely discussed, only a few researchers other than Witmer and Singer have attempted to measure presence and relate it to possible contributing factors. Barfield and Hendrix (1995) used simple, direct questions as measures of presence to show that update rate affects presence. (Update rate is the frequency (in frames per second) at which computer-generated images change in response to user actions or to other dynamic aspects of the simulation.) Prothero and Hoffman (1995) have shown that limiting the field of view near the eye, using an eye mask, reduces the amount of presence reported, again using a direct query about the subjective experience of presence. Furthermore, Slater, Steed, McCarthy, and Maringelli (1998) compared reports of presence with variations in visual stimuli (tree height in a virtual forest) and task complexity (counting deceased trees vs. counting trees and remembering location) and found positive associations between presence and the amount of participants' body movement.

Witmer and Singer (1998) have provided data that supports the concept of presence as a valid construct, as measured by the PQ. They have also shown both of the questionnaires to be internally consistent with high reliability (in earlier versions). Both the ITQ and PQ generate separate scales, derived by summing the responses to 7-point anchored Likert scales for different items. The ITQ scales were derived from previous research (on an earlier version using the same items, see Witmer & Singer, 1998). The ITQ has an *Involvement* scale reflecting participants self-reported tendency to become involved in different activities. There is also a *Focus* scale, relating the users tendency to maintain attention on current activities, and a *Games* scale, reflecting experience with video or computer games. An ITQ *Total* scale is generated by adding all items contained in these scales (without item repetition). The PQ scales include *Involvement & Control, Interface Quality, Naturalness, Auditory, Haptic,* and *Resolution* with a *Total* scale (also comprised of summed items). *Involvement & Control* items address how much the participant feels they had control and were involved in the experienced situation. *Interface Quality* addresses the perceived quality of the different interfaces used, whether they interfered with task performance or interrupted the experience. *Naturalness* addresses how natural the experience was perceived to be, and *Auditory, Haptics,* and *Resolution* address sound, physical manipulation, and visual acuity or capability.

The PQ scales have been shown to relate positively (although weakly) to task performance in VEs and to the ITQ scales, and are generally negatively related to simulator sickness as measured by the Simulator Sickness Questionnaire (SSQ) scales (Kennedy, Lane, Berbaum, & Lillienthal, 1993).

Although results relating measures of presence in VE to learning and performance in the VE and in the real world have been mixed (Bailey & Witmer, 1994; Witmer & Singer, 1994), many of the factors that appear to affect presence are known to enhance learning and performance (Witmer & Singer, 1998). Some situational factors that are believed to increase immersion, such as minimizing outside distractions and increasing active participation through perceived control over events in the environment, may also enhance learning and performance. Other factors may be more internal, such as tendencies toward involvement and selective attention, or familiarity with the task and situation. Some of these tendencies are independent of the situation (Witmer & Singer, 1998), and are measured with the ITQ. Therefore the ITQ should correlate positively and more highly with the initial PQ, obtained after the simplest VE situation. (As explained in the Methods section in the body of this report, during training participants' first VE experience was simple movement training.)

Because many of the factors involved in learning and performance logically should increase presence, it would be counter-intuitive if positive relationships between presence and performance, or between presence and equipment configurations that increase active participation, were not found. The ITQ and PQ have been administered before and after (respectively) many of the experiments conducted in the SSRU program. In our current experiment, results from the questionnaires were examined for relationships with the experimental variables, the VE equipment configuration, and the SSQ (Kennedy et al., 1993) results. The PQ was administered after several different phases in the experiment (see Table 1 and the Methods section in the body of the report). One expectation was that scores on the PQ would increase with any change in the VE that changes the amount of interaction required for minimal performance, or with increased proficiency based on practice. In this experiment, the initial training focused on learning to walk through the environment with a relatively normal body representation for position and orientation feedback. A later training session focused on equipment operation and team tasks (with an automated partner) with the same movement control. This later session should produce higher PQ scores than the earlier and simpler movement training, and will be tested using a planned comparison. This experiment also required repeated team missions, during which the teams were expected to improve in performance (learn to perform better on the tasks and with their team mate). The PQ was administered after the first and last of these missions, again with the expectation that increasing familiarity and capability would support increases in the experience of presence as measured by the PQ.

Methods

Participants

As described in the body of the report, participants were acquired at two geographical locations; Orlando, FL, USA, and Toronto, Canada. There were sixty-four training participants with a median age of 21. There were thirty-six participants assigned to teams for the mission rehearsal phase of the experiment, with a median age of 21.

Materials

As noted in the body of the report, all questionnaires were administered via computer using an Access[tm] database. The ITQ is replicated in Appendix D and the PQ is provided in Appendix E.

Procedures

The ITQ was only administered before the first VE session. The PQ was administered after the first VE session (movement training), the last training VE session (practice with an automated partner), the first team mission rehearsal and the last team mission rehearsal. Each time the PQ was administered the participants were instructed to answer the questions only based on the immediately preceding experience.

Results

Correlations were conducted between the ITQ scales and both the initial training PQ (response to the movement training VE, referred to as PQ1) and final training PQ (after VE task practice with an automated partner, referred to as PQ2) using the entire set of successfully trained participants for which all data were recorded. We used the Bonferroni adjustment on the traditional .05 alpha for a family of 28 comparisons to reduce the alpha level to .001 (see Tabachnick & Fidell, 1996). Because the statistical software (SPSS, Vs. 8.0) only generates p-values to the third decimal, this adjustment resulted in accepting any p-values of .001 or less as significant. The only significant correlations between the ITQ scales and PQ1 scales were between the ITQ Focus and PQ1 Total (r=.462, p<.001), PQ1 Involvement & Control (r=.401, p=.001), and PQ1 Resolution (r=.432, p<.001). There were no significant correlations between the ITQ and the PQ2 scales, using the same criteria and data set.

A series of planned comparison t-tests were conducted between the PQ scale scores for all trained participants over the two training PQ administrations. These analyses found significant differences between the PQ1 and PQ2 Total, Involvement & Control, Naturalness, Resolution, Auditory, and Haptics scores (see Table 1 for the t-values and p-values). The standard descriptive statistics for the administrations of these scales are also presented in Table

1. In every case but one, the second administration of the PQ resulted in higher scores for the scales. The only scale that did not significantly change was the Interface Quality scale.

Table 1
Presence Questionnaire Subscale Means and Standard Deviations from Training Administrations

Presence Questionnaire Scale	Initial Training		Final Training		*t*-Test (*N*=53)
	M	*SD*	*M*	*SD*	
Total	100.98	15.29	105.74	15.41	3.302 (*p*=.002)
Involvement & Control	61.34	8.37	63.83	9.01	2.854 (*p*=.006)
Interface Quality	16.77	3.16	16.11	3.15	1.45 (*p*>.05)
Naturalness	13.92	3.19	14.92	3.32	2.55 (*p*=.014)
Resolution	8.94	3.25	10.87	2.54	4.39 (*p*<.001)
Auditory	10.15	6.37	14.40	5.00	5.127 (*p*<.001)
Haptics	3.60	2.77	8.08	3.18	10.444 (*p*<.001)

Note: SPSS does not provide exact p*-values for those that are less than .001.*

In addition, the correlation between PQ1 scales and SSQ scales (also administered after the first VE session, see Appendix A) were investigated in order to help clarify that relationship. As with the ITQ correlations, a family of comparisons adjustment was applied that reduced the alpha level for significance to .001 (see above). PQ1 Total, Naturalness, and Involvement & Control scales correlated significantly with almost all of the SSQ scales, as shown in Table 2. The other PQ scales did not reach the adjusted level of significance with any of the SSQ scales.

The PQ scales were compared between the last VE training exposure (labeled PQ2) and the first team mission (PQ3) using planned comparisons. All analyses used the Bonferroni adjustment for the usual alpha level (.05) for the family of comparisons (yielding approximately .0071 for the individual comparison, see Tabachnik & Fidell, 1996). The analyses found significant differences between the PQ Total 2 and 3 (t (39) = 3.390, p = .002), and PQ Involvement & Control 2 and 3 (t (39) = 4.424, p < .001). Finally, planned comparisons were also conducted between the PQ administrations after the first (PQ3) and last (PQ4)

Table 2
Presence Questionnaire Subscale Correlations with Simulator Sickness Questionnaire Scales

Presence Questionnaire Scale	SSQ Total Severity	Nausea	OculoMotor Discomfort	Disorientation
Total	-.449 (*p*<.001)	-.397 (*p*=.001)	-.388 (*p*=.001)	-.411 (*p*=.001)
Involvement & Control	-.493 (*p*<.001)	-.457 (*p*<.001)	-.396 (*p*=.001)	-.440 (*p*<.001)
Naturalness	-.443 (*p*<.001)	-.364 (ns)	-.400 (*p*=.001)	-.429 (*p*<.001)

missions conducted by the teams. These analyses revealed significant differences between the PQ Total for 3 and 4 (t (42) = -3.367, p = .002) and PQ Involvement & Control 3 and 4 (t (42) = -3.262, p = .002). The standard descriptive statistics for the PQ scales from these administrations are also presented in Table 3.

Table 3
Presence Questionnaire Scales

Presence Questionnaire Scale	Final Mission Training (PQ2)		Initial Team Mission Rehearsal (PQ3)		Final Team Mission Rehearsal (PQ4)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Total	96.73	12.27	91.98	12.61	96.49	11.77
Involvement & Control	60.37	7.81	57.0	8.16	59.84	7.85
Natural	14.92	3.40	14.42	3.33	14.74	3.13
Auditory	14.18	5.09	15.0	3.47	14.86	3.50
Haptics	7.98	3.19	7.28	2.21	7.14	2.11

Discussion

The ITQ was correlated with the initial PQ responses, and not with the final training session PQ, although this was only the case for the ITQ Focus subscale. This seems to weakly support the argument that internal immersive tendencies would relate to PQ responses in an initial or simple immersive situation. The argument follows from the content of the ITQ Focus scale, which addresses tendencies toward attentional focus and the exclusion of extraneous or interrupting stimuli. The other scales address personal experience with different media and

interactive games. Better focus on the new experience would seem to be related to the experience of presence in the VE

There are three possible reasons for the increased PQ responses acquired after the last training session. The higher ratings could be based on the new task actions, which would immerse and involve participants by the novelty of the activity. The increased ratings could be based on the increased acceptance and skill acquired from multiple sessions. They might also arise from a direct comparison with the first experience in the VE configuration (simple movement). The participants were instructed to answer the subjective questions on the questionnaire based only on the immediately preceding activity, but humans are primarily comparative, as any student of human perception understands. There is no way to exclude possible direct comparison, but the time difference between questionnaire administrations and the instructions can be assumed to preclude comparative responses. Nothing about the VE configuration changed from the first training session to the last, with the exception of added tools and new tasks. The addition of interactive mechanisms (guns, grenades, sensors, and even door knobs that work) would seem to account for changes in Naturalness, Haptics, and Auditory scales. These are things that were not present in the initial movement training. The increased ability to interact with the environment would reasonably lead to an increase in Involvement & Control scores, but not necessarily in the Interface Quality scale. Obviously, the added contribution of items in the subscales would lead to higher values in the Total scale.

The PQ results dropped between the final training session and the first mission session, between one and six days later. The PQ scores then increased again after the last mission, which also followed the first in one to six days, to levels comparable to the final training session. The comparable intervals would seem to rule out the changes as a simple time function. It is not clear why the first mission session with a new human partner would be rated significantly lower than a partial mission segment with an automated partner. It does seem reasonable for the scores to increase from first mission to last. This could occur, without changes in the VE configuration or mission tasks, based on improving skills allowing greater immersion in the situation. This argument would seem to help explain the change between the last training and first mission presence scores. The changing task environment (increased difficulty) may have been sufficient to depress presence and involvement during the first mission, relative to the end of training. The participants would be, for the first time, interacting in a full mission situation with a new partner. Neither participant, although adequately trained, would be an expert in the mission roles and tasks. This performance difficulty would, theoretically, hinder the experiential flow. Increasing familiarity and proficiency would naturally lead to increased presence by the end of the eighth mission. Obviously, further experimental measures and manipulation would be required to verify the argument, but the findings do support the general conceptual underpinnings of the presence construct (Witmer & Singer, 1998).

The obvious next step in order to investigate the cause of the changes found during the training and mission segments of this experiment is to overtly manipulate the levels of task difficulty in order to show sensitivity of the PQ measure to task variables. An additional effort

might investigate the use of anchoring situations for a comparative measurement of presence, in search of additional sensitivity for measuring involvement and immersion.

Appendix B: References

Bailey, J. H., & Witmer, B. G. (1994). Learning and transfer of spatial knowledge in a virtual environment. *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting* (pp. 1158-1162). Santa Monica, CA: Human Factors and Ergonomics Society.

Barfield, W., & Hendrix, C. (1995). The effect of update rate on the sense of presence within virtual environments. *Virtual Reality: Research, Development, and Application, I*(1).

Hays, W. L. (1973). *Statistics for the social sciences*. New York: Holt, Rinehart, & Winston.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). A simulator sickness questionnaire (SSQ): A new method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3*(3), 203-220.

Prothero, J. D., & Hoffman, H. D. (1995). *Widening the field-of-view increases the sense of presence within immersive virtual environments* (Human Interface Technology Laboratory Technical Report R-95-4). Seattle, WA: University of Washington.

Slater, M., Steed, A., McCarthy, J., & Maringelli, F. (1998). The influence of body movement on subjective presence in virtual environments. *Human Factors, 40*(3), 469-477.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins College Publishers.

Witmer, B. G., & Singer, M. J. (1994). *Measuring immersion in virtual environments.* (Technical Report 1014, ADA 286 183). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Witmer, B. G., & Singer, M. J. (1998, June). Measuring presence in virtual environments: A presence questionnaire. *Presence, 7*(3), 225-240.

Appendix C. Participant Biographical Questionnaire

ID ___________

Please fill in the blank or circle the appropriate response.

1. What is your age? _____ years
(age)

2. What is your gender? female male
(sex) 1 2

3. Are you currently in your usual state of good fitness? yes no
(fitness) 1 0

4. How many hours sleep did you get last night? _____ hours
(sleep)

4a. Was it sufficient? yes no
(slepsuff) 1 0

5. Indicate all medications/substances you have used in the past 24 hours:
(medsubs)

CIRCLE ALL THAT APPLY

0 - none

1 - sedatives or tranquilzers

2 - aspirin, tylenol, other analgesics

3 - anti-histamines

4 - decongestants

5 - other (please list: ____)

6. Have you ever experienced motion or car sickness? yes no
(motsick) 1 0

7. How susceptible to motion or car sickness do you feel you are?
(motsscpt)

0	1	2	3	4	5	6	7
not susceptible	very mildly			average			very highly

8. Do you have a good sense of direction? yes no
(dirsnse) 1 0

9. How many hours per week do you use computers? _____ hours per week
(compuse)

10. My level of confidence in using computers is
(compcon)

1 2 3 4 5
low average high

11. I enjoy playing video games (home or arcade).
(vid_joy)

1 2 3 4 5
disagree unsure agree

12. I am _____ at playing video games.
(vid_con)

1 2 3 4 5
bad average good

13. How many hours per week do you play video games? _____ hours per week
(vidplay)

14. How many times in the last year have you experienced a virtual reality game or entertainment?
(vr_exp)

0 1 2 3 4 5 6 7 8 9 10 11 12+

15. Do you have a history of epilepsy or seizures? yes no
(epilepsy) 1 0

16. Do you have normal or corrected to normal 20/20 vision? yes no
(normvis) 1 0

17. Are you color blind? yes no
(colrblnd) 1 0

Appendix D

IMMERSIVE TENDENCIES QUESTIONNAIRE
(Witmer & Singer, Version 3.01, September 1996)

Indicate your preferred answer by marking an "X" in the appropriate box of the seven point scale. Please consider the entire scale when making your responses, as the intermediate levels may apply. For example, if your response is once or twice, the second box from the left should be marked. If your response is many times but not extremely often, then the sixth (or second box from the right) should be marked.

1. Do you easily become deeply involved in movies or TV dramas?

|_______|_______|_______|_______|_______|_______|_______|
NEVER OCCASIONALLY OFTEN

2. Do you ever become so involved in a television program or book that people have problems getting your attention?

|_______|_______|_______|_______|_______|_______|_______|
NEVER OCCASIONALLY OFTEN

3. How mentally alert do you feel at the present time?

|_______|_______|_______|_______|_______|_______|_______|
NOT ALERT MODERATELY FULLY ALERT

4. Do you ever become so involved in a movie that you are not aware of things happening around you?

|_______|_______|_______|_______|_______|_______|_______|
NEVER OCCASIONALLY OFTEN

5. How frequently do you find yourself closely identifying with the characters in a story line?

|_______|_______|_______|_______|_______|_______|_______|
NEVER OCCASIONALLY OFTEN

6. Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick and watching the screen?

|_______|_______|_______|_______|_______|_______|_______|
NEVER OCCASIONALLY OFTEN

7. What kind of books do you read most frequently? (CIRCLE ONE ITEM ONLY!)

Spy novels	Fantasies	Science fiction
Adventure novels	Romance novels	Historical novels
Westerns	Mysteries	Other fiction
Biographies	Autobiographies	Other non-fiction

8. How physically fit do you feel today?

|_______|_______|_______|_______|_______|_______|_______|

NOT FIT MODERATELY FIT EXTREMELY FIT

9. How good are you at blocking out external distractions when you are involved in something?

|_______|_______|_______|_______|_______|_______|_______|

NOT VERY GOOD SOMEWHAT GOOD VERY GOOD

10. When watching sports, do you ever become so involved in the game that you react as if you were one of the players?

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY OFTEN

11. Do you ever become so involved in a daydream that you are not aware of things happening around you?

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY OFTEN

12. Do you ever have dreams that are so real that you feel disoriented when you awake?

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY OFTEN

13. When playing sports, do you become so involved in the game that you lose track of time?

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY OFTEN

14. How well do you concentrate on enjoyable activities?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL MODERATELY WELL VERY WELL

15. How often do you play arcade or video games? (OFTEN should be taken to mean every day or every two days, on average.)

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY OFTEN

16. Have you ever gotten excited during a chase or fight scene on TV or in the movies?

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY OFTEN

17. Have you ever gotten scared by something happening on a TV show or in a movie?

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY OFTEN

18. Have you ever remained apprehensive or fearful long after watching a scary movie?

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY OFTEN

19. Do you ever become so involved in doing something that you lose all track of time?

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY OFTEN

20. On average, how many books do you read for enjoyment in a month?

|_______|_______|_______|_______|_______|_______|_______|

NONE ONE TWO THREE FOUR FIVE MORE

21. Do you ever get involved in projects or tasks, to the exclusion of other activities?

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY OFTEN

22. How easily can you switch attention from the activity in which you are currently involved to a new and completely different activity?

|_______|_______|_______|_______|_______|_______|_______|

NOT SO EASILY FAIRLY EASILY QUITE EASILY

23. How often do you try new restaurants or new foods when presented with the opportunity?

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY FREQUENTLY

24. How frequently do you volunteer to serve on committees, planning groups, or other civic or social groups?

|_______|_______|_______|_______|_______|_______|_______|

NEVER SOMETIMES FREQUENTLY

25. How often do you try new things or seek out new experiences?

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY OFTEN

26. Given the opportunity, would you travel to a country with a different culture and a different language?

|_______|_______|_______|_______|_______|_______|_______|

NEVER MAYBE ABSOLUTELY

27. Do you go on carnival rides or participate in other leisure activities (horse back riding, bungee jumping, snow skiing, water sports) for the excitment of thrills that they provide?

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY OFTEN

28. How well do you concentrate on disagreeable tasks?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL MODERATELY WELL VERY WELL

29. How often do you play games on computers?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL OCCASIONALLY FREQUENTLY

30. How many different video, computer, or arcade games have you become reasonably good at playing?

|_______|_______|_______|_______|_______|_______|_______|

NONE ONE TWO THREE FOUR FIVE SIX OR MORE

31. Have you ever felt completely caught up in an experience, aware of everything going on and completely open to all of it?

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY FREQUENTLY

32. Have you ever felt completely focused on something, so wrapped up in that one activity that nothing could distract you?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL OCCASIONALLY FREQUENTLY

33. How frequently do you get emotionally involved (angry, sad, or happy) in news stories that you see, read, or hear?

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY OFTEN

34. Are you easily distracted when involved in an activity or working on a task?

|_______|_______|_______|_______|_______|_______|_______|

NEVER OCCASIONALLY OFTEN

Appendix E

PRESENCE QUESTIONNAIRE
(Witmer & Singer, Vs. 3.0, Nov. 1994)

Characterize your experience in the environment, by marking an "X" in the appropriate box of the 7-point scale, in accordance with the question content and descriptive labels. Please consider the entire scale when making your responses, as the intermediate levels may apply. Answer the questions independently in the order that they appear. Do not skip questions or return to a previous question to change your answer.

WITH REGARD TO THE EXPERIENCED ENVIRONMENT

1. How much were you able to control events?

|________|________|________|________|________|________|________|

NOT AT ALL　　　　SOMEWHAT　　　　COMPLETELY

2. How responsive was the environment to actions that you initiated (or performed)?

|________|________|________|________|________|________|________|

NOT RESPONSIVE　　　　MODERATELY RESPONSIVE　　　　COMPLETELY RESPONSIVE

3. How natural did your interactions with the environment seem?

|________|________|________|________|________|________|________|

EXTREMELY ARTIFICIAL　　　　BORDERLINE　　　　COMPLETELY NATURAL

4. How much did the visual aspects of the environment involve you?

|________|________|________|________|________|________|________|

NOT AT ALL　　　　SOMEWHAT　　　　COMPLETELY

5. How much did the auditory aspects of the environment involve you?

|________|________|________|________|________|________|________|

NOT AT ALL　　　　SOMEWHAT　　　　COMPLETELY

6. How natural was the mechanism which controlled movement through the environment?

|_______|_______|_______|_______|_______|_______|_______|

EXTREMELY ARTIFICIAL	BORDERLINE	COMPLETELY NATURAL

7. How compelling was your sense of objects moving through space?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL	MODERATELY COMPELLING	VERY COMPELLING

8. How much did your experiences in the virtual environment seem consistent with your real world experiences?

|_______|_______|_______|_______|_______|_______|_______|

NOT CONSISTENT	MODERATELY CONSISTENT	VERY CONSISTENT

9. Were you able to anticipate what would happen next in response to the actions that you performed?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL	SOMEWHAT	COMPLETELY

10. How completely were you able to actively survey or search the environment using vision?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL	SOMEWHAT	COMPLETELY

11. How well could you identify sounds?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL	SOMEWHAT	COMPLETELY

12. How well could you localize sounds?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL	SOMEWHAT	COMPLETELY

13. How well could you actively survey or search the virtual environment using touch?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL — SOMEWHAT — COMPLETELY

14. How compelling was your sense of moving around inside the virtual environment?

|_______|_______|_______|_______|_______|_______|_______|

NOT COMPELLING — MODERATELY COMPELLING — VERY COMPELLING

15. How closely were you able to examine objects?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL — PRETTY CLOSELY — VERY CLOSELY

16. How well could you examine objects from multiple viewpoints?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL — SOMEWHAT — EXTENSIVELY

17. How well could you move or manipulate objects in the virtual environment?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL — SOMEWHAT — EXTENSIVELY

18. How involved were you in the virtual environment experience?

|_______|_______|_______|_______|_______|_______|_______|

NOT INVOLVED — MILDLY INVOLVED — COMPLETELY ENGROSSED

19. How much delay did you experience between your actions and expected outcomes?

|_______|_______|_______|_______|_______|_______|_______|

NO DELAYS — MODERATE DELAYS — LONG DELAYS

20. How quickly did you adjust to the virtual environment experience?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL — SLOWLY — LESS THAN ONE MINUTE

21. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?

|_______|_______|_______|_______|_______|_______|_______|

NOT PROFICIENT — REASONABLY PROFICIENT — VERY PROFICIENT

22. How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL — INTERFERED SOMEWHAT — PREVENTED TASK PERFORMANCE

23. How much did the control devices interfere with the performance of assigned tasks or with other activities?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL — INTERFERED SOMEWHAT — INTERFERED GREATLY

24. How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?

|_______|_______|_______|_______|_______|_______|_______|

NOT AT ALL — SOMEWHAT — COMPLETELY

25. How completely were your senses engaged in this experience?

|_______|_______|_______|_______|_______|_______|_______|

NOT ENGAGED — MILDLY ENGAGED — COMPLETELY ENGAGED

26. To what extent did events occurring outside the virtual environment distract from your experience in the virtual environment?

|________|________|________|________|________|________|________|

NOT AT ALL MODERATELY VERY MUCH

27. Overall, how much did you focus on using the display and control devices instead of the virtual experience and experimental tasks?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT VERY MUCH

28. Were you involved in the experimental task to the extent that you lost track of time?

|________|________|________|________|________|________|________|

NOT AT ALL SOMEWHAT COMPLETELY

29. How easy was it to identify objects through physical interaction; like touching an object, walking over a surface, or bumping into a wall or object?

|________|________|________|________|________|________|________|

IMPOSSIBLE MODERATELY DIFFICULT VERY EASY

30. Were there moments during the virtual environment experience when you felt completely focused on the task or environment?

|________|________|________|________|________|________|________|

NONE OCCASIONALLY FREQUENTLY

31. How easily did you adjust to the control devices used to interact with the virtual environment?

|________|________|________|________|________|________|________|

DIFFICULT MODERATE EASILY

32. Was the information provided through different senses in the virtual environment (e.g., vision, hearing, touch) consistent?

|________|________|________|________|________|________|________|

NOT CONSISTENT SOMEWHAT CONSISTENT VERY CONSISTENT

Appendix F. Simulator Sickness Questionnaire (SSQ)

Adapted from Kennedy, Lane, Berbaum, & Lilienthal (1993)

ID ________________ Date

Instructions: Please indicate how you feel right now in the following areas, by circling the word that applies.

1.	General Discomfort	None	Slight	Moderate	Severe
2.	Fatigue	None	Slight	Moderate	Severe
3.	Headache	None	Slight	Moderate	Severe
4.	Eye Strain	None	Slight	Moderate	Severe
5.	Difficulty Focusing	None	Slight	Moderate	Severe
6.	Increased Salivation	None	Slight	Moderate	Severe
7.	Sweating	None	Slight	Moderate	Severe
8.	Nausea	None	Slight	Moderate	Severe
9.	Difficulty Concentrating	None	Slight	Moderate	Severe
10.	Fullness of Head	None	Slight	Moderate	Severe
11.	Blurred vision	None	Slight	Moderate	Severe
12.	Dizzy (Eyes Open)	None	Slight	Moderate	Severe
13.	Dizzy (Eyes Closed)	None	Slight	Moderate	Severe
14.	Vertigo*	None	Slight	Moderate	Severe
15.	Stomach Awareness**	None	Slight	Moderate	Severe
16.	Burping	None	Slight	Moderate	Severe

*Vertigo is a disordered state in which the person or his/her surroundings seem to whirl dizzily: giddiness

** Stomach awareness is usually used to indicate a feeling of discomfort which is just short of nausea.

ARE THERE ANY OTHER SYMPTOMS you are experiencing right now? If so, please describe the symptom(s) and rate its/their severity below. Use the other side if necessary.